

Clinton's true colours on trade?

There is a danger that President Bill Clinton, in seeking better terms of trade with Japan, will impoverish the rest of us without improving the lot of the United States.

Mrs Margaret (now Lady) Thatcher, in her time, was fond of comparing national economies with those of ordinary households. Her presence might have been valuable at the meeting last week in Washington between President Bill Clinton and Mr Kiichi Miyazawa, the prime minister of Japan. Her homily might have gone like this. Nations are like households, right? Suppose that one of the largest, say **A**, is living so far beyond its means that it has to borrow money from its friends to cover its domestic expenses. Suppose, further, that **A**'s economic connections with its friends — **B**, **C**, **D** and the like — consist only of the sale among themselves of goods and services, perhaps the equivalent of a sack of vegetables for a pair of shoes. **A** (but also **B**, **C**, **D** and the others) then have a problem. If the quantity of goods and services on offer and the quantity of money in circulation remain fixed, **A**'s extravagance means that its debt to its friends grows year by year, while its friends cannot afford to buy everything that **A** would like to sell them.

The analogy should now be transparent. **A** is the United States, whose extravagance is most easily measured by the federal budget deficit. **B**, **C**, and **D** are the trading partners of the United States (but principally Japan), which finance the US deficit by capital transfers, either by buying US federal securities or, equivalently, by building factories or by buying real estate. Naturally, this willingness to help the United States live beyond its means is not sheer altruism: the lenders calculate that they will gain more in the long run from their investment than they lose in the short run from the consumption they forgo. So there is a trade imbalance as well. If there were not, overseas lenders would not have the funds to lend, and the whole burden of the US deficit would fall on US taxpayers, who would then have less to spend on purchases.

Evidently, from last week's proceedings, Clinton does not trust this simple arithmetic. He would like to make the trade imbalances go away, believing that US factories would then be working full-time again, turning out products for Japan and the other rich countries of the world. Whence the suggestions last week in Washington that the principles of the bilateral agreement on computer chips reached six years ago should be extended to other commodities. If not in as many words, Clinton is asking for a prearranged share for the United States in the Japanese domestic market in a variety of goods and services. Should not trade be 'fair' is what he asks? That would be correct if he simply means that the Japanese government should not stand between Japanese people and the goods and services US (or other) enterprises wish to sell

them. But a quota system as Clinton seems to have in mind would impoverish us all, the United States included.

The Clinton recipe for managed trade, if successful and if extended beyond Japan, would lead to an entirely unexpected result — to a trading system indistinguishable from the discredited Soviet organization called Comecon, abandoned soon after Mr Mikhail Gorbachev's accession to power in 1986. That offered producers throughout the then Soviet bloc guaranteed outlets for their (often unwanted) goods. Comecon was one of the chief reasons why Soviet technology lagged behind that in the West for decades on end. It would be fanciful to suppose that Clinton seeks to recreate in the industrialized West an incubus like that, but that is the direction in which his policies are pointing.

But how else should a powerful country seek to rid itself of its trade deficit? The Thatcherian homily points to one way; if the budget deficit were made to vanish, as Clinton creditably avows it will, there would be no problem. (Dollars would be worth more yen, so that even the present volume of trade would be more nearly in balance.) Trade deficits would no doubt still exist, but they would be simply a measure of countries' relative productive efficiency and would be adjusted over time by movements in the value of different currencies.

But what is to happen in the many years before the US budget deficit melts away? Luckily, the caveats accompanying the household homily are not absolutes. The wealth to share among households is not fixed for all time, but can be increased. Technical change is one agent. International trade is another. And the two are linked. As the costs of development increase, international markets are essential to the commercial success of innovations. And innovations by advanced economies are essential means of vacating roles that can then be filled by countries now just on the edge of industrial manufacturing. It may be difficult to tell all this to people in places such as the United States who are still out of work, but Clinton should try. Bullying Miyazawa may be easier, and makes for better headlines, but is mistaken. □

Reactors and compromise

President Bill Clinton's budget compromises are sending conflicting signals about his seriousness.

PRESIDENTS of the United States have no time to say where they stand on all the issues for which they are responsible.



That goes without saying. If President Bill Clinton were required to make a public speech on, say, his policy on the National Science Foundation, his plans for the more effective management of the Department of Energy, his long-term policy for the huge urban centres of America and the future role of nuclear power in the United States, he would have no time to decide what to do about Bosnia, for Mr Boris Yeltsin and for the 35 million people not covered by health insurance in the United States. So the president can mostly only "send signals". Everybody understands. But some of Clinton's signal-sending in the past few weeks has been disturbing for the sense of casualness it displays.

Take, for example, the case of the National Science Foundation (NSF), supposed only a few weeks ago to be an essential ingredient of the plan to "put America back to work". Just how was never explained, but never mind. But then the President runs into trouble with his political opponents, who are against his \$19 billion plan (either on the grounds that it is too little or too much). So he cuts his request for funds to Congress almost in half (see page 688). It sounds fair, of course. NSF will probably get something extra, but Clinton's complaisance has undervalued the agency's reputation as an agent of technical change. If half will do as well as a considered whole, why not settle for nothing?

The President's decision about the plutonium-burning reactor at the Argonne National Laboratory's Idaho site (once the National Reactor Testing Center) is less defensible (see page 683). Its budget for next year has been reduced by five-sixths from \$125 million. When, a few decades from now, the history of nuclear power in the United States is written, people will be mystified that the budget has been so abruptly trimmed, but still not reduced to zero. What, they will be asking, was meant to be the signal? That the Clinton administration was hostile to nuclear power as such, but worried about jobs? Or anxious to keep its options open, but on a scale that would not give offence?

The technical advantages of Argonne's experimental nuclear reactor are plain: it is a model for a breeder reactor that consumes as fuel the long-lived components of spent reactor fuel, the heavier actinides as well as plutonium. As such, it is potentially also a means of getting rid of material that might be used for making weapons. Operating such a novel machine is not child's play. Whatever the designers say, the safety of the equipment must be tested thoroughly before larger versions are built, whenever that may be. If the Idaho project is not a long-term project, it is nothing.

Nuclear power has necessarily now become a long-term business. Dreams that endless energy might already be generated from uranium have been defeated by carelessness in the United States, the Soviet Union and elsewhere. But that state of affairs cannot last. It would be strange if a world unable to burn coal for fear of global warming nevertheless turned its back on the most important source of energy that does not emit carbon dioxide. That is when historians will regret that there is no practical experience of how to operate reactors that consume the long-lived waste from other reactors. Thanks to Clinton's apparently boundless capacity for compromise, it may then be too late. □

Life is short

Eastern Germany should tolerate the temporary indignities of unifications for the sake of the better times ahead.

It is natural enough that people in eastern Germany should be disappointed that many of the plum jobs at newly invigorated research institutes in the six eastern *Länder* should have gone to applicants from what was West Germany. There may be even more substantial cause for grumbling that the flow of relatively junior people from west to east has fallen so far below the target of 10 per cent advised by the Wissenschaftsrat in 1991. But before grumbling turns sour, it is to be hoped that people in eastern Germany will acknowledge that what may seem like a takeover of eastern institutions by carpetbaggers from the west is a manifestation of exceptional circumstances that will soon, with luck, themselves melt away.

The survey described on page 685 provides the first objective evidence of how posts in reorganized research institutes are being filled. Conducted by telephone by *Nature's* Munich office, the survey provides merely a snapshot of what has happened so far in the staffing of the newly organized institutes. Much may change as more posts are filled. But it is also, in any case, inevitable that the past isolation of the old German Democratic Republic (GDR) from all but the old Soviet bloc will have weighed against the appointment of easterners to senior research posts; their bibliographies will have been thin on publications in recognized journals, their experience of the big wide world will often have been scant, while their experience of the management of research will usually have been different in character from that habitual in the rest of Germany.

Taken together, those pleadings may seem like a justification for bias of a kind. So in a sense they are. But the bias is inevitable, a simple product of the deprivations (as seen from the west) of life in the GDR. After all, has not the great dismaying discovery of the rush to unification been that all kinds of people in the East, from shopkeepers to university professors, have been disadvantaged by the past. A more substantial cause for complaint (if one is really needed) is that, as yet, too little has been done in Bonn to regularize the positions of those in eastern academic institutions who have been accused of political complicity with the old regime, and who have lost their jobs in consequence. There is some justice in Honecker's plaint, at his abortive trial earlier this year, that he should not be tried for following the orders of a then legal government recognized as such in the west.

The best hope, for 'Ossies' and 'Wessies' alike, is that the appointments that have been made to senior positions at eastern institutes are indeed as excellent as they could have been, so that the institutes concerned will prosper. One important objective of the whole exercise, apart from yielding some excellent science, is to assist in the cultural integration of the two parts of Germany. Long before those newly appointed to eastern posts come to retire from work, it is to be hoped that the distinction between 'Ossie' and 'Wessie' will have disappeared. □

US fast-breeder reactor fights for survival after Clinton trims budget

Washington. This week the US Congress will lift the curtain on what could prove to be a long, hot summer of conflict over whether to close the country's last major experimental nuclear reactor.

Energy secretary Hazel O'Leary is likely to be questioned about the Integral Fast Reactor (IFR) at the Argonne (Illinois) National Laboratory when she testifies today (Thursday, 22 April) before a House of Representatives appropriations subcommittee. The Clinton administration originally sought to terminate the project, being carried out at Argonne's Idaho site. But after strong lobbying from local congressmen, the president agreed to request \$20 million — down from a current budget of \$125 million — for continued research into the recycling of the class of heavy-metal elements known as actinides (plutonium and other fissile elements).

This compromise, however, appears to have satisfied no one. In the months ahead, supporters will request sufficient funds to keep the reactor running and prepare it for commercial application, while opponents will demand closure.

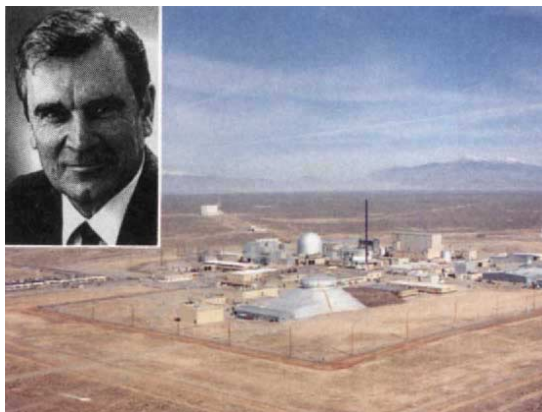
The winding down of the reactor is the main component in the Department of Energy's plan to reduce spending on its nuclear programme from \$345 million to \$284 million in the 1994 fiscal year, which begins on 1 October. But laboratory officials say that the high cost of terminating the project means that little money will be saved.

The chairman of the House appropriations subcommittee, Tom Bevill (Democrat, Alabama) supports the reactor and is likely to put additional money into his funding bill. His counterpart in the Senate, J. Bennett Johnston (Democrat, Louisiana), has also supported the project. When the legislation moves to the floor of each house, the vote will test the strength of the pro- and anti-nuclear energy lobbies for the first time since Congress voted ten years ago to shut down the Clinch River fast breeder reactor project in Tennessee.

The IFR programme employs 1,000 people at Idaho and another 500 at Argonne's main site in Illinois. Supporters say that the reactor design is "intrinsically safe", with no possibility of a meltdown because expansion of the fuel upon overheating will moderate the reaction and lower the temperature. They say the design could not only generate cheap power safely, but would also recycle existing nuclear waste from generators and weapons, producing small volumes of waste with a relatively short half-life of 200 years.

"This is potentially the biggest energy

source bar none", says Argonne's director of engineering, Chuck Till, who has led the project since its inception in 1984. He points out that the amount being spent on IFR — around \$120 million a year — is "down by



The Integral Fast Reactor in the Idaho desert and project director Chuck Till.

a factor of ten" from the budget for fast-reactor research a decade ago.

Representative Richard Durbin (Democrat, Illinois), a supporter of the IFR programme, says it will be "an uphill battle" to save the reactor. "Those unalterably opposed to nuclear power will never favour it, but if you believe that nuclear power can be used responsibly, you will support it", he says.

As with other 'big-science' programmes, the potential for collaboration with the Japanese is being cited by both opponents and proponents of the IFR. The Department of Energy would like to attract a foreign sponsor — almost certainly Japan — to help support its proposed actinide recycling pro-

gramme. But cuts in federal support are likely to repel, rather than attract, overseas investment.

The environmental lobby does not accept the claims being made on behalf of IFR and is concerned that the administration may back away from campaign promises to abandon research on nuclear power. Scott Denman, director of the Safe Energy Communication Council, fears that the administration's early concessions on actinide recycling mean it has lost the will to kill the reactor. "When a programme moves in this way it sends a message that it will continue", he says. "They've spent \$700 million on this since 1984, and I'm worried that they'll find ways of keeping it going."

The debate on the reactor involves not just its safety or its impact on the environment but also its implications for nuclear proliferation. Supporters of the reactor say that its output of plutonium, neptunium and americium is of no military value and that it can thus be sold to countries, ranging from Japan to Ukraine, that need plants to consume existing nuclear wastes. Their opponents say that any country buying a fast reactor could adapt it to create fissile material for military purposes.

The last time the project was challenged in Congress — last summer in a vote called by since-retired Michigan congressman Howard Wolpe — it was endorsed by a 2:1 margin. This year brings a new administration and more than a hundred new members unfamiliar with the IFR but united in their professed hostility to the financing of (other people's) pet projects. For the Integral Fast Reactor at Argonne, that hostility could prove fatal.

Colin Macilwain

France to close four observatories

Paris. France has decided to close its four national observatories to meet the growing costs of its contributions to international facilities. The Pic-du-Midi observatory will be closed in 1998, with the others — Nancy, Haute-Provence and Côte d'Azur — staying open at least until the end of the decade with a smaller staff and budget as a postponement but not a reprieve.

Closure of Pic-du-Midi will save FF13.6 million (US\$2.5 million) a year, with reductions in the others amounting to another FF50 million. The growing importance of international facilities justifies the loss of the smaller national observatories, accord-

ing to the Institut National des Sciences de l'Univers (INSU), which controls spending on large astronomical observatories. France is paying FF300 million towards construction of the Very Large Telescope in Chile as a member of the European Southern Observatory and FF87 million to build a solar telescope THEMIS with the Italians in Tenerife, Canary Islands. It also spends FF17 million a year to operate a telescope on Mauna Kea and FF31.4 million for a submillimetre radioastronomy telescope built with Germany and Spain. The institute's annual budget has remained static at around FF1,200 million. **Declan Butler**

Japan reviews first guidelines for gene therapy trials

Tokyo. Japan last week laid the foundations for a system to regulate gene therapy by releasing guidelines compiled by a scientific committee of the Ministry of Health and Welfare. But clinical trials (see below) are not imminent: the report avoids assigning authority for final approval of gene therapy experiments, and nothing can be done until the Ministry of Education, Science and Culture, which is responsible for university hospitals, clarifies its position on use of the technique.

As expected, the guidelines prohibit its use on reproductive cells because of the impact on future generations, and they say the technique should be applied only against terminal illnesses such as brain tumours, AIDS and cancer for which no effective remedy exists. They call for review of protocols by internal review boards in hospitals and medical research institutes. The boards will be overseen by the ministry which, in turn, will be advised by a committee of specialists similar to the Recombinant DNA Advisory Committee of the US National Institutes of Health. But the committee's authority remains unclear.

Some Japanese newspapers and scientists have interpreted the guidelines as delegating most decision-making to the individual internal review boards and to the head of the hospital or institute. Kunio Yagi, director of the privately funded Institute of Applied Biochemistry in Gifu Prefecture, says that he and researchers at Nagoya University have already asked the chairman of the university's internal review board, a former student of Yagi's, to examine their preliminary plans for clinical trials on a technique for introducing interferon genes into brain cells to treat brain tumours.

Yagi hopes to complete experiments on nude mice within three months and to win approval for human clinical trials "within

this year". Once he has obtained the board's approval, he does not anticipate difficulty in clearing the central government committee "provided we have the consent of the patient and our technique is scientifically acceptable and safe".

But Masaaki Terada, director of the National Cancer Centre Research Institute and one of the members of the committee that wrote the guidelines, anticipates a more comprehensive review. "It is my understanding that all gene therapy experiments will be discussed by the committee of specialists after [they have been reviewed by] the institutional committees", he says.

Masanori Fukushima of Aichi Cancer Centre, a vocal critic of the medical research system, believes that the guidelines on informed consent are flawed because they stipulate that consent can be obtained from patients orally. Without written forms such as are required in the United States, he says, it will be impossible for review committees to know if consent was really obtained.

The Ministry of Education, Science and Culture, which oversees Japan's universities, including Nagoya University, has yet to make its position clear. Yagi says the education ministry is "quite relaxed" about the issue and that it will leave the decision to individual universities. But others expect the committee of specialists to play an important role because of the small number of scientists in Japan who are knowledgeable about gene therapy and the ethical, social and legal issues raised by the new techniques.

The guidelines point to a need to win public understanding of gene therapy and the importance of conducting business openly, but they offer little advice on how this can be done. Meetings of academic and government committees are typically conducted in private.

David Swinbanks

NIH to support embryo bank for mice and rabbits

Washington. Researchers would gain access to the increasing number of genetically altered strains at reasonable cost and without restrictions on their use under a plan by the US National Institutes of Health (NIH) to create a repository of frozen embryos of genetically altered rodents and rabbits.

NIH has allocated \$1 million for the repository, an accompanying inventory and another facility, possibly in a separate location, that would develop new and improved cryo-preservation techniques, particularly for freezing mouse sperm.

Grant applications are being sought from commercial breeders and non-profit enterprises, colleges, medical centres and universities with the necessary in-house expertise. Requests must be limited to not more than \$350,000 in direct costs for the first year.

"We're not trying to develop something from scratch here", says Lance Liotta, deputy director for intramural research at NIH. The facility would preserve the flood of newly discovered strains that have no commercial value but may be of interest to researchers outside the laboratories in which they were developed. At present, the task of maintaining and distributing these strains rests with the researcher.

Judy Vaitukaitis, director of the National Center for Research Resources, which is providing the funding, says that the repository would provide researchers with frozen embryos or animals "probably at less than cost". Although researchers funded by NIH will be given priority, the repository will be open to researchers with other sources of research support.

The initiative is a "step in the right direction", says Janet Rossant of the Samuel Lunenfeld Research Institute in Toronto, Canada, who is concerned about its emphasis on cryopreservation as a means of maintaining the resource. Cryopreservation is "only part of the story", she says, and is unlikely to replace the need to keep strains that are in continuous demand (which would include some of the so-called 'gene knock-out' mice) as active breeding colonies. Having to breed sufficient animals for an experiment from frozen embryos can build in a delay of six months or more into a research programme, she says, and not all laboratories, if supplied with frozen embryos, would have the technical knowledge to perform the necessary embryo transfer procedures.

The deadline for proposals is 23 June, with the awards to be made after two rounds of scientific review before the close of the current fiscal year, which ends on 30 September.

Diane Gershon

AIDS patients are first in line

Tokyo. The first clinical trials of gene therapy in Japan are likely to involve AIDS patients and to be carried out by a private company. Mitsuru Miyata, editor of the newsletter *Nikkei Biotechnology*, predicts that the blood product manufacturer Midori Juji (Green Cross Corporation) will be first on the basis of the worldwide marketing rights it holds for a gene therapy treatment for AIDS developed by Viagene Inc. of San Diego, California.

Japan's Ministry of Health and Welfare, which regulates the pharmaceutical industry, is concerned about the safety of retrovirus vectors used in gene therapy because of the risk of cancer. But Viagene's vector system, carrying the full-length copy of the HIV envelope protein, has already been approved in the United States, a fact that is expected to ease its path through the Japanese ministry. Funds for developing the gene therapy guidelines came from the ministry's anti-AIDS campaign, and thus it seems likely that the first clinical trials will involve AIDS patients.

D.S.

West Germans are taking most top jobs at eastern institutes

Berlin and Munich. An attempt to 'seed' new research institutes in eastern Germany with western scientists is occurring only at the top, according to a *Nature* survey of 97 institutions.

In 1991, following the closure of all institutes of the Academy of Sciences in the former East Germany, Germany's science council, the Wissenschaftsrat, recommended that their successors should be populated by a mixture of eastern and western scientists. A figure of about 10 per cent westerners — working at all levels — was considered to be an ideal target to speed up cultural integration.

But that has not happened. Although statistics are not collected, the government acknowledges its failure to achieve such a balance, the result being too many westerners at the top and almost none at the bottom.

With a substandard research infrastructure and problems such as chronic shortage of housing and poor schooling, the eastern *Länder* are not very attractive to post-doctoral researchers used to better standards. But the chance for promotion changes the equation. The opportunity to head a new institute or department has proved to be a powerful magnet for those facing obstacles to career advancement in the west.

According to the *Nature* survey, the open competition in the new *Länder* has resulted in the selection — generally by western scientists — of a majority of their colleagues to top positions in institutes of basic research. Only one of nine heads of the new national research centres is an east German, and roughly half the department heads and three-quarters of group leaders hail from the west.

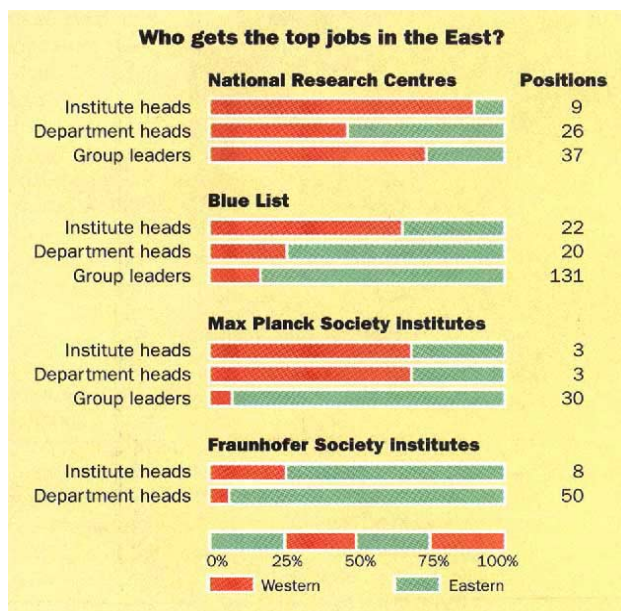
Similarly, only one of the three heads so far recruited for the newly founded Max Planck institutes are east Germans. Although staffing of the new institutes is not complete, department heads follow the same pattern; by contrast, the vast majority of group leaders appointed so far are from the east.

So-called 'blue list' institutes — independent research establishments with equal funding from the federal and *Länder* governments — have proven to be a little more receptive to east Germans. Although also

very much in flux, one third of the institute heads so far recruited, and around 80 per cent of the department heads and group leaders, are east Germans.

The applied research (Fraunhofer) institutes, by contrast, have recruited predominantly from the east, a reflection of the relative strength of product-orientated research in the former DDR German Democratic Republic and possibly also the fact that researchers in this area were less likely to be politically compromised.

Martin Holzhauer, a biochemist from



East Berlin, says that the figures are disappointing, particularly when so many research groups were given such positive reviews by the Wissenschaftsrat and so many genuinely good east Germans have been overlooked. Detlev Ganten, head of the Max Delbrück Centre in Berlin-Buch, agrees. "The depressing part for east Germans is that the top positions are being taken up by west German scientists — this must be considered against a background of 100 per cent east Germans at the outset. It is an important psychological point in the whole process of reunification."

One of the eight national research centre heads in the new *Länder* who hailed originally from the west, Ganten believes that local scientists will fare better over the next few years. Although the overall ratio of westerners to easterners in his centre is less than one in ten, Ganten hopes to appoint more group leaders and department heads from within to strike a better balance at the upper levels.

Alison Abbott & Cornelia Koob

CHINA IN BRIEF

Beijing. The Chinese government has granted 100 research institutions the right to make their own deals on foreign trade in a move intended to increase exports of high-technology products. High-technology products comprise less than 5 per cent of the total, and the government has given institutes operated by 28 ministries the freedom to promote their own technologies in the hope of doubling that figure. Such exports have grown by an average of 32 per cent a year since 1986, led by 3,000 high-technology companies operating in 52 national high-technology development zones.

■ Determined not to be left behind, a hundred Chinese universities have joined to form a high-technology group corporation that will occupy a high-rise office building under construction in Shanghai as well as five factories to manufacture products developed in university laboratories. The State Education Commission encouraged the creation of the corporation to hasten the transfer of new technology from campus to market. Chinese universities have had mixed success in forming individual companies based on academic research, with Beijing and Qinghua universities leading the way, and there remains concern that the quality of undergraduate education will suffer if too much attention is given to applied work.

■ The government has proposed a restructuring of ministries to reduce employment by 20 per cent and to hasten the development of new technology. The ministry of aviation and aerospace industry, for example, will be converted into two specialized corporations, one for aviation and one for aerospace, to encourage the rapidly growing aerospace industry. At the same time, the ministry of machinery and electronic industry will become separate ministries to reduce the administrative burdens on the burgeoning electronics industry.

■ China's desire to cooperate with the rest of the world to stimulate economic development must remain secondary to the country's need to defend itself, warned the government's minister of national security in a recent televised address explaining the duties of each citizen under a law passed in February. The minister, Jia Chun-Wang, said that "a very small number of foreign hostile forces have never given up hope of penetrating into China and are doing everything they can to overthrow the Chinese government", and science and technology are viewed as major targets. The budget and details of a high-technology programme begun in 1986, for example (see *Nature* 359, 177; 1992), are regarded as state secrets and remain classified, along with most of the meteorological, topographical and surveying data collected by government scientists.

You Qui Li

NIH takes Stewart and Feder off the misconduct beat

Washington. The US National Institutes of Health (NIH) have reassigned two controversial scientists and told them to end a decade of work investigating scientific misconduct.

The scientists, Ned Feder and Walter Stewart, hope to overturn a decision by their bosses at the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) to abolish their two-man biophysical histology laboratory. But it is not clear how hard their supporters, including Representative John Dingell (Democrat, Michigan), who chairs the committee that oversees NIH, will fight to reinstate them. The men have survived previous attempts by NIH to constrain their activities, but their latest foray into allegations of plagiarism against historian Stephen Oates may have been the last straw.

Feder and Stewart became part of a growing debate during the 1980s on the nature and extent of scientific misconduct by analysing the involvement of co-authors on fraudulent papers written by John Darsee during three years at Emory and Harvard universities. The paper was published after a three-year odyssey through the scientific press (*Nature* 325, 207; 1987), and the scientists soon became a magnet for those with allegations of scientific impropriety. Their most controversial role was as consultants to a four-year investigation by Dingell's oversight and investigations subcommittee into the contents of a 1986 *Cell* paper, of which Nobel laureate David Baltimore was one of the co-authors. The government's final report on the case is pending and may be issued this summer.

On 9 April, the two men were told to turn over their misconduct files to the Office of Research Integrity (ORI), formed last summer as a successor to the Office of Scientific Integrity to investigate allegations of scientific misconduct by those with grants from NIH or its sister agencies within the US Public Health Service. Feder has been reassigned with effect from 10 May to a job reviewing funding proposals from outside scientists, and Stewart is to be transferred to a biophysical chemistry laboratory within the institute.

The reassignment is being made for three reasons, according to a press spokeswoman for the institute. "History is not part of our mission", says Elizabeth Singer of NIDDK. "The Oates affair takes them well beyond biomedical research." In addition, she says that the creation of ORI to investigate mis-

conduct "makes it unnecessary for anyone to go to Wally and Ned" with their complaints. Finally, NIDDK must lose 30 posts in the next year as its part of a government-wide slimming down by the Clinton administration, and the biophysical histology laboratory was deemed expendable.

Both men say that they will fight to retain their current positions and to continue their work on misconduct. "The problem isn't

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Stewart, left, and Feder at NIH.

Oates", says Feder. "It's scientists like Walter and me who want to take a hard look at misconduct. That's what NIH wants to stop." The men have long claimed that, as tenured NIH staff, their licence to investigate scientific misconduct stems from the same type of intellectual freedom enjoyed by their academic colleagues. In contrast, NIDDK officials seem to be drawing the line at work that falls outside the scope of the institute. NIH director Bernadine Healy was informed of the institute's decision but has no comment on the matter, according to Anne Thomas, an NIH spokeswoman.

Although the men are not at present working on any case for Dingell's Energy and Commerce committee, a committee source says that their reassignment "raises troubling questions that we would like to have answered". Those include the precise reasons for the move — the letter notifying them of their new positions does not mention the Oates case — and whether NIH intends to have someone else carry out their work on misconduct. "ORI's role is to investigate complaints and consider sanctions against those found guilty of misconduct", says the committee source. "We're interested in knowing if there will be anybody undertaking the type of scholarly research on misconduct that Stewart and Feder have been doing."

Jeffrey Mervis

Promega extends fight with Roche over Taq enzyme

Washington. Promega Corporation of Madison, Wisconsin, is challenging the validity of a key patent held by Hoffmann-La Roche of Nutley, New Jersey, for the heat-stable enzyme *Taq* DNA polymerase — the enzyme that drives the polymerase chain reaction (PCR).

A victory by Promega, who filed papers last week in US district court in New Jersey, would loosen the grip that Roche and its exclusive licensee, Perkin-Elmer Corporation of Norwalk, Connecticut, now has on the sale and pricing of the enzyme when it is used in PCR. Promega and six other companies have licences from Roche to sell *Taq* polymerase for less commercially lucrative purposes such as DNA sequencing.

The patent for *Taq* polymerase was issued to scientists at Cetus Corporation of Emeryville, California, in 1989 and purchased by Roche in 1991 when Cetus sold its PCR business to Roche for \$300 million. It covers the purified enzyme *Taq* DNA polymerase either isolated from the bacterium *Thermus aquaticus* or derived from a genetically engineered organism containing a gene that codes for the production of the enzyme.

Last autumn, Roche sued Promega as part of its attempt to clamp down on those thought to be infringing on its patent rights. Roche claims that Promega is violating a 1990 licensing agreement between the two companies under which Promega is permitted to manufacture and sell *Taq* polymerase as a stand-alone enzyme for purposes other than PCR.

Promega has refused to discuss the basis for its challenge, but Randall Dimond, chief technology officer for Promega, says that his company has found evidence in the literature that *Taq* polymerase was purified from the bacterium *T. aquaticus* before Cetus became involved. In particular, he cites a paper by four Russian scientists published in *Biokhimiya* (45, 644–651; 1980) and reprinted that year in *Biochemistry* that was entitled "Isolation and properties of DNA polymerase from extremely thermophilic bacterium *Thermus aquaticus* YT1".

Dimond believes that Roche sued his company because it is concerned that people are buying the enzyme from cheaper sources than Perkin-Elmer and using it in PCR. Promega is the second largest supplier of *Taq* polymerase worldwide (behind Perkin-Elmer) and its prices are 50 per cent lower than those of Perkin-Elmer. Christine Aylward, a spokeswoman for Roche Molecular Systems Inc., says that Promega is challenging Roche's *Taq* patent to divert attention from the first lawsuit.

Diane Gershon

Britain rediscovers industry-academia LINKs

London. The Department of Trade and Industry (DTI) has lifted a temporary freeze on the LINK programme, the government-wide initiative that encourages industry to undertake collaborative research with universities and other publicly funded research institutions. The programme is expected to receive an enthusiastic endorsement in the government's forthcoming White Paper on science and technology, now destined for publication early in June.

LINK has had a chequered career since its formation in 1986. Although the principle of strategic research projects funded equally by government and industry is widely accepted, the programme was slowed by bureaucratic rules and complex review procedures. But new rules introduced in 1991 following a critical review of its progress by the consulting firm of Segal Quince Wicksteed in Cambridge (see *Nature*, **351**, 431; 1991) have made LINK into one of the Conservative government's most widely-quoted success stories. There are now more than 300 individual projects covering 30 programmes, ranging from protein engineering to (most recently) the biological treatment of soil and water.

Last autumn, however, the department put a temporary freeze on new initiatives and began a review of all research programmes after it experienced budgetary problems in the wake of the collapse in value of the pound and Britain's departure from the European Exchange Rate Mechanism. With respect to LINK, the DTI investigated concerns from industry that the results of individual projects did not meet the needs of the market place.

The review is expected to lead to more precise definitions of the goals on each project. And funding for LINK is likely to be held level for the near future after several years of rapid expansion.

Even so, the DTI believes that LINK provides both a politically acceptable and cost-effective way of bridging the gap between industry and the academic community by giving academic scientists the incentive and opportunity to work on problems of long-term interest to industry. It has also persuaded companies in industries not regarded as high-technology — for example, waste management or cake-making — of the value of collaboration with university research teams.

The DTI's decision pleases officials at the Science and Engineering Research Council, which has been the department's main partner in many of the individual LINK programmes. It has also been welcomed by Robert Malpas, chairman of the Cookson Group and of the LINK steering committee, which was concerned that the DTI's funding difficulties might damage LINK.

Malpas argues that LINK, which he de-

scribes as a "runaway success", has allowed academic institutions and small and large companies to develop a common language and purpose. He would like to see the programme move into a second phase, identifying neglected areas of strategic research and

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

This LINK project studies the three-dimensional structure of proteins.

using government and private funds to prime the pump by sponsoring experiments in these areas.

Any such development would tie in with plans of the Office of Science and Technology (OST) to sketch out the anticipated areas of growth for high-technology industries. That information would help researchers to design initiatives to be funded through the research councils and government departments. Speaking last month at a meeting in London, William Waldegrave, the minister responsible for science, extolled a tech-

nique known as 'technological foresight' as a way of building a national consensus around areas of research and technology that would most benefit Britain.

OST officials, remembering that previous public backing for key technologies produced commercial failures such as the Concorde, emphasize that they are not trying to pick winners but are rather, according to Waldegrave, trying to generate a "shared perspective" between industry, the government and the academic community. A preliminary study of information storage technology has already demonstrated the potential effectiveness of such an approach.

Some are hoping that the promise of a full-blown assessment could become a central feature of the White Paper. But various government departments are cautious about the value of such an exercise and are concerned that an attempt by the OST to produce a single list of 'critical technologies' or generic technologies could be unduly influenced by the views of individual companies or universities.

The most likely outcome at present is a compromise under which the OST will use technological forecasting to help scientists to judge for themselves what types of research are the most likely to coincide with the longer-term interests of industry. That would leave individual government departments free to draw up their own lists and preserve OST's overall responsibility to target new areas of growth.

LINK, now responsible for "identifying generic science and technologies of wide applications and commercial relevance", already seeks to achieve such a goal. One option for Waldegrave is to bring LINK under the aegis of the OST and thus the cabinet office, a recommendation that the government rejected two years ago.

David Dickson

Cabibbo to lead pontifical science academy

Munich. Nicola Cabibbo, an Italian particle physicist, has been appointed president of the Vatican's Pontifical Academy of Sciences, a body of 80 or so international scientists who organize scientific conferences on ethical issues such as transplants, the use of global resources and even — at the risk of rummaging in God's territory — cosmology. Their conclusions are summarized for use by the pope's office.

Cabibbo, professor of theoretical physics at the Second University of Rome, was formerly head of Italy's nuclear physics institute and is a delegate to the CERN, European Laboratory for Particle Physics in Geneva. He discovered the phenomenon of current and quark mixing in weak interactions. He was elected as leader of the pontifi-

cal academy for a four-year period directly by Pope John Paul II. The academy selects its own members, and membership is not limited to Catholics.

The pontifical academy, refounded in 1976, considers itself the direct descendant of the seventeenth century Accademia dei Lincei, which was itself often disbanded and recreated and which counted Galileo as a member. The secular Italian National Academy of Sciences claims the same lineage.



Cabibbo

Alison Abbott

NSF search advances as budget retreats

Washington. Three women — two of whom are active scientists and one a college president — are said to be the leading candidates to succeed Walter Massey as director of the US National Science Foundation (NSF). The rumours of a pending announcement came as President Bill Clinton last week withdrew nearly half of his request for a \$207 million increase this year for NSF that was predicated on its ability to generate a short-term boost in the US economy. It is part of a \$19.5-billion economic stimulus package that Clinton has pared by \$4 billion in an attempt to appease Republicans in the US Senate.

Those on the 'short list' being considered by presidential advisers are Linda Wilson, president of Radcliffe College in Cambridge, Massachusetts; Rita Colwell, head of the Maryland Biotechnology Institute in Baltimore; and Sandra Faber, professor of astronomy at the University of California, Santa Cruz. Wilson has refused several requests for an interview, while Colwell and Faber declined to answer whether they were being considered.

Colwell and Faber, along with others on a longer, earlier list, believe that the next director will be important in deciding whether NSF assumes a larger role in addressing strategic needs while retaining its primary mission to support academic research. Last summer, the US Congress urged NSF to spend a larger part of its \$2.7-billion

budget in that way, and last autumn a commission appointed by Massey issued a controversial report, encouraging NSF to do both and to foster closer ties between industry and university researchers.

In today's climate of fiscal austerity, the balance between the two becomes paramount. About 55 per cent of NSF's supplementary budget request — and a similar amount of its overall budget — is spent on six areas that are seen as helping to strengthen the US economy, from advanced manufacturing to biotechnology, or that address national goals, such as high-performance computing or understanding global climate change. Massey thinks such a balance "is about right". But he left last month, two years into a six-year term, to become provost of the University of California system, and a new director will have some freedom to choose which course to follow.

Other issues are equally pressing. One is how to pay for large and costly instrumentation and facilities — from telescopes to ocean-going research vessels — without cutting too deeply into the number and size of grants to individual investigators, some of whom rely on such equipment for their work. "The cost of new facilities is the critical problem facing astronomy", says Faber, "and I really don't know if there is a way out."

A related problem is the growing use by universities of lobbyists to win funding for

new laboratories and buildings not requested by the funding agency, the so-called 'pork-barrel' projects. Says Faber, "the key to a happy community is to ensure that the procedures for approval are above reproach".

Another problem is the rising number of investigators competing for grants and the resulting decline in funding rates — now less than one in four in some fields. This factor competes for attention with the increasing cost of doing research, and together they could more than swallow up whatever additional funding is obtained, leaving little room for new initiatives. For example, NSF has promised to use almost half of whatever extra funding it receives this year to increase the size of grants rather than to make new awards, but each year the NSF director faces pressure on both fronts.

The job has considerable appeal. As head of an independent agency, the NSF director does not have to deal with the same suffocating bureaucracy that hinders the National Institutes of Health, for example, and potentially he or she has greater access to the White House. In addition, the six-year term offers more job security than most presidential appointments. "It's a job that any US scientist or engineer would find interesting", says Donald Langenburg, chancellor of the University of Maryland and a former deputy NSF director who is thought to be a strong contender if the president decides not to appoint a woman.

Jeffrey Mervis

DNA testing helps British bring better pig to market

London. A leaner and happier pig is the latest exhibit in an attempt by Britain's Ministry of Agriculture, Fisheries and Food to persuade the public to accept the application of molecular genetics to animal breeding.

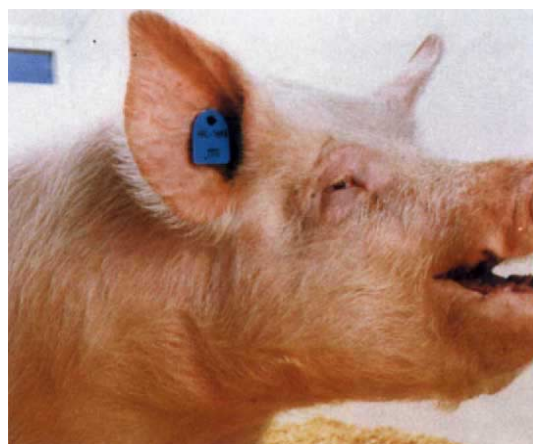
The new pig is guaranteed to be free of the 'halothane' gene, a genetic mutation that traditionally could be detected solely through exposure to the anaesthetic vapour halothane. The mutation is linked to increased leanness in meat and has become increasingly common as pigs are bred for this characteristic. But it has also been found to reduce the capacity of meat to hold water (an undesirable trait in keeping meat moist in shops) and to increase the animal's susceptibility to stress.

The gene was identified two years ago by scientists at the University of Toronto in Canada. The mutation itself was already in the process of being bred out in a programme carried out by the Cotswold Pig Development Company Ltd that began in the early 1980s. However, since the halothane test works only for individuals in which the mutation is expressed — and not those that are merely carriers — the ultimate success of

the project relied on the development of a DNA detection kit based on the Canadian discovery.

The resulting pig, which is guaranteed to be free of the mutation, is said by Cotswold to be the first application of DNA technology in farm animals. The company believes that an important advantage, besides producing pigs that are more profitable for farmers and whose meat is more attractive in supermarkets, is a reduced level of stress for the pig itself. "We should be proud that science can deliver this kind of advantage, and not be frightened of it", says John Gummer, the agriculture minister.

Wary of potential protests from animal rights group, the company is quick to point out that the development of the new pig has not required the use of genetic engineering. Animal welfare groups admit that they are impressed by the elimination of a genetic trait which makes the pig less prone to stress but that it has not reduced their distrust of genetic engineering more



Cotswold boar with a tag in its ear declaring it to be free of the mutation.

generally. "We are still concerned about the wider applications of these new techniques, since they put a foot on the accelerator of our ability to produce new varieties of animals purely for commercial reasons", says Martin Potter of the Royal Society for the Prevention of Cruelty to Animals.

David Dickson

Theology and science

SIR — I write to correct the errors in the article about the Starbridge Lectureship in Theology and Natural Science (*Nature* 362, 380; 1993).

(1) The University of Cambridge went to no "lengths" to attract my endowment. No one from the fund-raising office approached me before I made my decision. For years there had been a campaign by distinguished academics in the fields of theology and natural science to establish a teaching post in this area. My decision to fund the lectureship was made in the light of this knowledge; I knew I would have the appropriate academic backing right from the start. No academic ever approached me for funds.

(2) I did not decide in 1980 to go down the "well-trodden road to celebrity status". My religious conversion, a gradual event, began in 1983 and I said nothing about it to anyone for several years. In 1991 I wrote an article on the subject, but only because the editor of the *Church Times* asked me to do so; he thought the mechanics of the experience would be of interest to his readers. The article had nothing to do with publicizing *Glittering Images*, which by that time had been in print for over three years.

(3) I did not spend six years away from writing. It was about fifteen months. (May I, in passing, make it quite clear that the title of my novel *The Wheel of Fortune* was not based on anything written by Richard Dawkins? It was taken from the philosophical writings of Boethius.)

(4) I did not have a "school-leaver's certificate in religious studies". I have A-levels in English literature, Latin and history, and an upper-second class honours degree in law.

(5) It is an error to imply that my statement, which you quote, about the complementarity of science and theology is my opinion alone. I have been studying theology for ten years but I do not pretend to be an original thinker. On this point I am merely repeating the view of scientists and theologians who are far more distinguished in their fields than I am.

(6) It is an error to imply that all my novels can be described as 'blockbusters' (publishing slang for a long novel that sells well). My career has had three phases: my first six novels were short, my next five were long and my last five have been, by the standards of today, of average length. Only the middle five can be accurately described as 'blockbusters', and these had nothing to do with the endowment of the lectureship.

(7) It is an error to state that my novel *Glittering Images* funded the Starbridge

lectureship. Royalties from this book were donated to another charitable cause.

(8) It is an error to imply that the novels that do fund the lectureship are so devoid of merit that their earnings "pollute" the University of Cambridge. The series has been well received by the literary critics of (among others) the *Sunday Times*, *The Observer*, the *New York Times* and the *Washington Post*. The fourth novel in the series won a literary prize. As the result of the series I have been invited to speak or lecture at (among others) the Universities of Cambridge, Oxford and Durham.

May I conclude this letter by saying how sorry I am that you failed to do the proper research into my life and work before printing this article. Surely it is not in the best tradition of science to be careless about the truth and negligent in the matter of research? I would have expected such an esteemed publication to show more intellectual rigour.

Susan Howatch

Flat 2,
42 Great Smith Street,
London SW1P 3BU, UK

SIR — As an agnostic, I am happy to acknowledge the very great contributions to science of Christian thinkers. Newton, Pascal, Faraday, Maxwell, who brought such a lively imagination to the physical sciences, and who taught us how to investigate and rationalize the occult forces of nature, were all in their own way devout religious men.

In the life sciences, it was the anti-Christians who took over last century, and then went on to dominate the scientific culture this century. Unfortunately, their influence spread so much wider, with Haeckel becoming the revered prophet of the Nazis, and Darwin the champion of the New Right.

Our moderns, such as Richard Dawkins and John Maddox, not only oppose the religious view on principle, but do their best to deride and censor any secular studies that may have remote religious overtones. Don't waste time on these matters, they say. Put the money into genetics, into drugs, into transplants, into suppression of the immune system, into HIV. Give it to the Gallos of this world. Whether we can survive much more of this experimentation, there is good reason to doubt.

Cambridge is, I trust, still a liberal university, and can appreciate Susan Howatch's contributions to our society. The obnoxious sneers from *Nature*, although quite predictable, demean the whole scientific community.

My hope is that the money may be

used to help broaden scientific education in the university, with suitable historical and philosophical content to give students an understanding of how we travelled from Pythagoras and Plato to the present day. It is surely not sufficient that they only solve equations, or read DNA sequences. With serious attention given to the evolution of ideas, they may perhaps avoid some of the tragedies of this century.

John Evans

81 Cherry Hinton Road,
Cambridge CB1 4BS, UK

SIR — Your reaction to the report of the Howatch donation to Cambridge was not, I suppose, unexpected, but is I submit entirely mistaken. I was very glad to hear of the gift.

I have passed most of my working life as a professor of chemistry, and at the same time a member of the Anglican Church, for a considerable period churchwarden at my local church. I have been able to reconcile the two to my own satisfaction, and not in any of the ways that you mention. I am much more conventional than you would suppose, both as a scientist and as a churchman. It would be unproductive to go into this here, suffice it to quote the well-known remark that nuclear physics may prove to be not only more strange than we suppose, but more strange than we are capable of supposing.

What right has anyone to assume that, because they cannot imagine any way in which science and religion can co-exist, God cannot do so?

The declining moral values of modern society are, or should be, a cause of concern to us all. Moral values in the past were based on religion, and if, as those of us who are Christians would claim, they depend on religion, then the perceived discrepancy between science and religion is of great importance, and certainly worth investigation.

If study convinces the scholars that there is no antagonism between the two views of the world, this must be good for the moral climate. If study convinces the scholars that the gulf is unbridgeable, then I do not know what would happen, principally because I do not believe in this outcome. But if this were to be the case, it would be important to know it, to comfort the materialists and help them to prepare their view of the world, and good luck to them too.

If, as of course is most likely, the scholars come to no conclusion, they are not likely to have done any harm, and may shed much illumination on a complex, confused, and above all, important question.

David A. H. Taylor

12 Avenue Road,
Scarborough YO12 5JX, UK

SIR — Why do you say that “the University of Cambridge should have thought harder before accepting a donation joining science and theology” when this is for a lectureship in divinity, one of the longest established faculties in the university?

Perhaps because of doubts that there can be any useful relationship between the natural and the theological sciences, or whether theology is a sensible subject for study nowadays anyway. Science can tell us approximately when the Universe is supposed to have begun and why it developed as it did, and when life originated on Earth and how it has evolved to where we are today. These are scientific questions which theology cannot and should not attempt to answer.

But what science cannot tell us is *why* the Universe started at that time, nor indeed why it exists at all. Some may say that because such questions are unanswerable they ought not to be asked. But, granted that science cannot provide the answers, the questions still remain. Theology does try to answer them, and that the answers may not always be thought to be the right ones doesn't mean that this isn't a proper subject for rigorously academic study.

C. B. Goodhart

Gonville & Caius College,
Cambridge CB2 1TA, UK

SIR — In your attack on the establishment of a lectureship in theology and natural science in the University of Cambridge, you make the extraordinary claim that the sciences can be pursued with academic rigour, whereas the study of the relationship between science and theology necessarily lacks such rigour. Apart from being an unnecessary slur on the faculties of theology which form an integral part of many universities, such comments also ignore the obvious fact that many of the leading figures in the history of science have displayed considerable intellectual energy in the task of relating their science to their faith. A list of such individuals would certainly include Kepler, Galileo, Boyle, Descartes, Bacon, Pascal, Ray, Newton, Priestley, Kelvin, Faraday and Maxwell, as well as a large number of twentieth century scientists. If anyone thinks that investigation of the interactions between science and theology lacks academic rigour, then a recent book by John Brooke, editor of the *British Journal for the History of Science (Science and Religion — Some Historical Perspectives)*, CUP, 1991, should dispel such doubts.

Today, modern science is placing enormous power in the hands of humanity which can be used for either creative or destructive purposes. There is an urgent moral responsibility to harness such power for the common good and

the fulfilment of that responsibility is becoming increasingly complex. Rather than the outmoded attempt to draw sharp boundary lines around their fields of investigation, there is a need for a more humble approach whereby scientists welcome attempts to relate their knowledge to theology and moral philosophy. The decision to establish the lectureship in theology and natural science is therefore not only a continuation of a tradition of partnership between science and theology which goes back many centuries, but also provides a further valuable opportunity to explore the ethical responsibilities of the scientific community with respect to the applications of their science.

R. J. Berry (University College, London, UK);

D. C. Burke* (University of East Anglia, Norwich NR4 7TJ, UK);

J. N. Hawthorne (University of Nottingham, UK);

R. B. Heap (Babraham Institute, Cambridge, UK);

John Houghton (Rutherford-Appleton Laboratory, UK);

C. Russell (Open University, UK).

* For correspondence

Two-way adaptors

SIR — V. Bauchau¹ referred to the contribution of Dover² dismissing Dawkin's claim for Universal Darwinism³. As to the fact that laws may deal with very general statements rather than with precise predictions on a system's detailed development, I agree with Bauchau completely. Whether, however, evolution by natural selection is such a universal law, as he called it, will depend on the units of selection he may have in mind. If one goes high enough in the general hierarchy of organisms, then, of course, evolution is a zero-sum-game between winners and losers. But as soon as forms of cooperation and function sharing come up, one can no longer separate the losers from the winners, that is, from those who have been selected⁴.

Another point is that it is not compelling to say that the (social, organic or physical) environment will exclusively define the conditions to which the organism has to adapt in order to survive. One can also say that organisms through their own evolution have defined the criteria to be met by the environment, and that in all cases where the environment is not appropriately adapted, the organism will try to modify the environment according to these criteria by means of acting in the wider sense (including emigration). Then, as Waddington⁵ said, the organism will select the environment rather than the reverse. In Darwinian parlance, a species will die out (and be replaced by others) if its members and the environment are ill-matched. Yet the same can

happen also to a special environment itself if it is occupied by species which developed tools to modify their surroundings. One may well continue to speak in terms of conditions to which organisms have to adapt and to use the notion of selection accordingly, but then it has to be made clear that, to a great extent, organisms themselves have brought about these conditions as well as they can modify them.

Olaf Diettrich

Commission of the European Community,
DG XII E-1, Biotechnology,
Rue de la Loi 200,
B-1049 Brussels, Belgium

1. Bauchau, V. *Nature* **361**, 489 (1993).

2. Dover, G. A. *Nature* **360**, 505 (1992).

3. Dawkins, R. *Nature* **360**, 25–26 (1992).

4. Diettrich, O. *Evol. Cogn.* **2**, 163–188 (1992).

5. Waddington, C. H. *Nature* **182**, 1634–1638 (1959).

University funding

SIR — Your recent Opinion article on “University competition” (*Nature* **362**, 2; 1993) expounded the view that the research assessment exercise has to a considerable extent followed the Matthew principle “to him who hath shall be given”. To satisfy the requirements of including some reward for research merit, of providing safety-net funding as well as of keeping within the available budget, the Higher Education Funding Council resorted to a formula incorporating a crude capping mechanism. As a result, the University of Cambridge, which came top of the list, received an increase of only 15 per cent in its research award rather than the figure of nearly 30 per cent which it might otherwise have been given. This means in practice that a grade 5 department at Cambridge will receive about 10 per cent less research funding under the award than a similar department at an uncapped university.

High-quality research departments have been further penalized by two additional factors. First, the research assessment scales terminated at 5 rather than at 7 or even 10, which might have been a better indicator of true international quality. Second, early experience of the recent overhead transfer exercise suggests that this will cause a substantial loss of support money which the more research-intensive departments will find impossible to cover by any other means.

I do not believe that we are complacent about any of the features of the new system — either its wider challenges to which you rightly draw attention, or its defects which we hope will swiftly be rectified.

Archie Howle

Cavendish Laboratory,
Cambridge CB3 0HE, UK

Sunshine for light in the night

Neville Williams, Ken Jacobson and Harold Burris

People in rural regions of the developing world are hungry for electricity to meet the most rudimentary of needs. Small solar-power systems have proved themselves to be an appropriate technology.

FOR two billion people — 70 per cent of the population of the developing world¹ — the productive day ends when the Sun goes down. Each evening, millions of families retreat into smoky rooms lit only by kerosene lamps or wood fires. The lot of this vast proportion of the Earth's population would and could be improved by the provision of a resource so many elsewhere take for granted. That resource is electricity.

Our view, supported by practical experience, is that photovoltaic solar-power systems are the answer, or at least one answer. Such small solar-power systems (SSPS) provide a clean and inexhaustible source of electricity^{2,3}, and can be designed to service a single household. They consist of a photovoltaic solar module which charges a storage battery, like that used in motor cars, which stores the electricity ready for use (see box). The overwhelming initial need, as public opinion surveys in India have indicated¹, is for domestic lighting and broadcast reception. But as expectations grow and economics dictate, SSPS can be used more widely for refrigeration, for heating both water and air, and for small business activities.

What are the alternatives? Families in rural parts of the developing world are usually beyond existing electrical grids and are not attractive commercial targets for utilities. They do not offer the economies of scale found in servicing consumers in urban areas, nor do they provide big industrial loads to warrant extension of main power lines. Dams and hydroelectric power, once seen as highly promising, also have serious limitations. Such schemes usually mean the flooding of fertile valleys and the displacement of large numbers of people. During dry seasons and droughts, reservoir levels drop and power output may be restricted. To build dams, developing countries often incur huge debts which are serviced, in part, by sales of rural cash crops. It was originally thought that hydroelectric power would help make rural populations more productive, but as it has turned out investment in rural grid extension has never been justified economically⁴.

Solar-powered residential electric lighting and communications systems are already available⁴⁻⁶ and are truly environmentally friendly. About 100,000 families, farms and businesses in the

developing world are currently using them. Complete photovoltaic modules can be imported; alternatively the solar cells are imported and the modules are assembled locally. As demonstrated in Sri Lanka, Kenya, the Dominican Republic and Zimbabwe, locally made rubber-cased car batteries, with their thick plates, work well as energy-storage devices for this application and can last for more than 5 years. Complete systems cost from US\$350 to \$750, including shipping, tariffs and installations, and should last for 20 years. Typically, nearly 60% of the initial cost is in the solar module itself, another 25% or so in the battery and lights, 10% in wiring and circuitry, and 5% in installation costs. (However, the recurrent cost of battery replacement can be higher than the original cost of the module over its 20-year life span.)

People in rural areas throughout the world have demonstrated a willingness to pay for electricity, and to purchase it from anyone who can provide it. In the Dominican Republic⁵, a revolving loan programme managed by a local solar-energy development association has provided financing for nearly 2,000 household installations. Photovoltaic companies linked with rural cooperatives have financed more than 10,000 systems in Indonesia. In Sri Lanka, commercial

banks offer 5-year consumer loans to "Light up your home with solar power".

Also in Sri Lanka, small private companies are providing SSPS for individual families. Suntec modules and electrical systems are manufactured locally by the company Power & Sun. With loans from two national development banks, Power & Sun has demonstrated that home solar systems can address the needs of the 70% of Sri Lankans who do not have access to the main power grid or local generators. This company first established a nationwide network of dealers (usually small appliance shops), trained a cadre of self-employed technicians and installation technicians, and educated its market through advertising and promotion.

Today, Singer Sri Lanka Ltd, the appliance retailer, also distributes Suntec systems and provides instalment financing. In addition, several model international donor programmes are trying to make solar rural electrification a widespread reality. BP Solar introduced 1,000 home photovoltaic lighting systems in a Sri Lankan village last year as part of an Australian aid project. The panels were manufactured by Power & Sun, under licence, using BP solar cells. The Ministry of Energy is now advocating solar electrification as the cheapest option for energy supply. ►

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Wired for light, sound and vision in rural Sri Lanka. The 9-watt lamp and the radio and television are all powered from a battery charged by a small solar panel on the roof.

Small nonprofit organizations are also involved. Enersol Associates in Cambridge, Massachusetts, has worked primarily in the Dominican Republic and Central America, while the Solar Electric Light Fund in Washington, DC, founded by one of us (N.W.), is performing a similar role in Asia and Africa. The Solar Electric Light Fund has set up two revolving loan schemes in Sri Lanka, and more recently has arranged to replicate these programmes in western China; there, 200 million rural people are without electricity.

The Sao Paulo Declaration, from a UN-sponsored photovoltaic workshop held in September 1991, calls for one million photovoltaic home-lighting systems to be installed by 1995 and ten million by the year 2000. Declarations are one thing, however, realizing them another, and there are several barriers to the dissemination of this technology^{4,7}. First, there is the ability of rural populations to afford the relatively high initial cost. New types of financing schemes are needed and additional central sources of capital will have to be identified to provide the backing for local financing. Happily the worldwide awareness of global warming and other environmental problems associated with large hydro-electric projects and conventional fossil-fuelled power plants should allow a healthy slice of energy-sector money to be diverted to environmentally benign solar programmes. The World Bank is considering incorporating solar rural electrification into energy lending programmes for India, Sri Lanka and Indonesia.

Because solar electric lighting systems are cost-effective, it is possible to develop applications so that the loans given to purchase the systems can be repaid. Drip irrigation systems, small power tools, sewing machines, office equipment and other light electrical loads can be profitable if the appropriate engineering is done to couple solar electric power to the equipment. Profitable applications leading to quick repayment will allow revolving loan funds to be distributed to larger groups.

A further key to making SSPS available to many people who could not afford the down payments or monthly payments is for public utilities to purchase large numbers of units. Roof- or pole-mounted units would then be leased to rural customers at a nominal monthly rate. Utilities could obtain cheaper financing and components for the systems than can individual consumers or small cooperatives. This approach will

Photovoltaic-based SSPS

Solar cells convert light to direct current electricity. Light impinging on single crystal or polycrystalline silicon wafers generates hole-electron pairs which are separated at a p-n junction fabricated into the wafer. The separation of charge by the junction creates a potential difference that produces a current in an external circuit. The current is used to charge a battery placed in the circuit and the battery, in turn, provides electrical power, on demand, for lights and other appliances. Individual solar cells are encapsulated and fabricated into a multicell unit, the solar panel.

A standard roof- or pole-mounted solar module, with typically 36 10-cm-diameter solar cells charging a car lead-acid battery, provides enough power for both household fluorescent lighting and a radio and television. For example, in the tropics, the average daily insolation will often exceed 5 kWh m⁻². Employing a 40 Wp solar module (a module producing 40 watts at noon under direct sunlight), this insolation will produce enough electricity for 3 fluorescent lights each operating for 3 hours, as well as 3 hours of radio and 3 hours of television. The solar modules are generally mounted on roof tops or on poles which allow them to be turned to maximize daily charge near the Equator and to keep them cool. Car battery storage provides up to 5 days of power supply during overcast periods. Locally assembled battery charge indicators and charge-discharge controllers are the key to long battery life. When users are well-trained on how to keep their batteries alive by not allowing them to over-discharge, they become experts at managing their power consumption. High efficiency fluorescent d.c. light fixtures using electro-polished aluminium reflectors and inverter ballasts minimize lighting power requirements. For example, one 9-watt fluorescent tube produces as much light as a 50-watt a.c. incandescent bulb.

be tried as part of a Global Environment Facility photovoltaic lighting project underway in Zimbabwe; the project has a budget of \$7 million.

In addition, there is a pressing need to develop even smaller and cheaper systems. For most rural families, the cost of a 50 watt SSPS is still too high, even if credit can be arranged. In this respect, smaller systems such as the 'solar lantern', which use a smaller solar module and a built-in storage battery to provide power for one good light and a radio, hold high promise. The aim is to produce rugged, durable and efficient systems for less than \$100.

Other impediments to the spread of SSPS are that rural populations need to be aware of this technology and to develop a trust in its capabilities. In some places, power-delivery institutions have created a self-serving 'myth of electrification', which has it that eventually everyone will be reached by the main power grid. This is not so, and it discourages new investment in alternative technology. Other problems include high tariffs on imported components of SSPS, unexpected national economic difficulties and the lack of adequate after-sales service in rural areas.

As incomes increase and more efficient solar modules become a reality, expandable SSPS with more power output will be required to meet more sophisticated requirements. The technical

challenge will be to design devices and appliances that cool, heat and cook, and which are highly efficient in their use of solar power.

Larger systems will be feasible, especially if better methods of energy storage can be devised. This is one of the big challenges. New methods of regulating both charging and discharging promise to lengthen battery life⁸. Imported, long-life, deep-discharge batteries would be more desirable, but at present they are too expensive for most rural households. More efficient capacitive storage batteries will also be useful if the cost can be kept low. Eventually, it may be that other power-storage methods such as flywheel energy accumulators, pumped storage or solar-hydrogen energy systems will provide alternatives to batteries. But — and it cannot be over-emphasized — cost is the crucial consideration. Unless the improvements can be manufactured cheaply and in high volume, the people most in need of them will not be able to afford them.

For now, however, the success of thousands of small solar photovoltaic electric systems in the field shows that this technology works as it is. Power needs are rapidly increasing and, according to World Bank estimates, meeting the demand for electricity in the developing countries will cost up to \$1 trillion in the 1990s. Only one-fifth of this amount has been identified¹. Even rudimentary SSPS, we believe, can go some way towards filling that gap. □

Neville Williams is at the Solar Electric Light Fund, 1739 Connecticut Avenue NW, Washington, DC 20009, USA. Ken Jacobson is in the Department of Cell Biology and Anatomy, University of North Carolina, Chapel Hill, North Carolina 27599, USA. Harold Burris is the Global Environment Facility photovoltaic project consultant, United Nations Development Programme, Harare, Zimbabwe.

1. Lovejoy, D. *Natural Resources Forum* May, 102–110 (1992).
2. Zweifel, K. *Harnessing Solar Power* (Plenum, New York, 1990).
3. Hubbard, H. M. *Science* **244**, 297–304 (1989).
4. Hankins, M. *Solar Rural Electrification in The Developing World. Four Country Case Studies: Dominican Republic, Kenya, Sri Lanka and Zimbabwe* (Solar Electric Light Fund, Washington, DC, 1993).
5. Hansen, R. D. & Martin, J. G. *Natural Resources Forum* May, 129–144 (1988).
6. Williams, N. *Solar Today* Nov./Dec., 22–24 (1991).
7. Annan, R. H., Malbranche, P. & Hurry, S. *Strategy for Disseminating/Commercializing Proven Renewable Energy Technologies*, 1–13 (United Nations Development Programme, June, 1991).
8. Harrington, S. R. *Charge Controller Technology*, BOS Workshop, SOLTECH 92, Photovoltaic Systems Research (Sandia National Laboratories, 1992).

Uncertainty in quantum theory

Protestations to the contrary notwithstanding, people have been busy bridging the gap between deterministic wave mechanics and the notion that everything is uncertain; quantum mechanics is the better for it.

It is said that when Schrödinger devised his quantum wave equation in the turbulent year of 1926, he believed for a little while that he had shown that the familiar simplicities of classical mechanics had, after all, been restored. And that would be understandable. For the Schrödinger equation is a perfectly deterministic equation exactly comparable to the equation of motion of a classical mechanical system. If you know the state of a system (represented by a wavefunction $\psi(x, t)$, a function of position and time respectively) and enough about its properties to define the equation of motion, you can set out to calculate the state of the same system at any later time.

One of the reasons for the general misunderstanding of quantum mechanics (but there are several, see *Nature* **361**, 493; 11 February 1993) is the mismatch between the determinism embodied in Schrödinger's equation and the general opinion that quantum mechanics is all about uncertainty. That, of course, is how it happened historically. It fell to Bohr to argue for the statistical interpretation of the different quantities calculated by Schrödinger on the one hand and Heisenberg on the other. Thus, having solved a wave equation, Bohr advocated that $\psi(x, t)^* \psi(x, t)$ should be taken as the square of the probability distribution of the position coordinates x at time t . (The quantity $\psi(x, t)^*$ represents the complex conjugate of the regular wave function.)

The awkwardness of this arrangement is that it puts two characteristic features of quantum mechanics into two quite different intellectual pockets. The business of calculating the motion of a system is as matter-of-fact as even a Lord Kelvin could ask, but then the result must be interpreted as if the stochastic nature of microscopic physics is a kind of afterthought. It is no wonder that over the past half century there have been endless complaints from those who allow that Schrödinger's equation "works", but who cannot (or, at least, could not) accept that physics is inherently uncertain.

To be fair, it is aesthetically (to put it mildly) unsatisfactory that a theory should not embody the means by which it is to be interpreted. Newton's laws meet that test: space and time are absolute, and the quantities calculated from the newtonian equations of motion, positions, velocities and the like, have an obvious interpretation. The same is true of special relativity, except that the quantities are now relative. But what, in 1926, could be said for a theory producing

a well-defined result (a wavefunction changing with time) which is then interpreted as a probability with hardly better justification than the question, "What else can it mean?"

The argument was still raging in the 1950s. We forget that, then, the Soviet view of quantum mechanics was that it was an "idealist" doctrine, a term of particular abuse in the lexicon of the times. It was also generally overlooked that the question had for practical purposes been settled by the late Richard Feynman, whose particular way of doing quantum mechanics incorporated uncertainty in a successful scheme of calculation.

What is the chance that a particle will go from A to B ? Why, simply list all the paths between the two points (even those not classically allowed) and calculate the likelihood that the particle will make each journey, making proper allowance for the phase (from which interference phenomena spring). What Feynman had done was to take literally Bohr's argument about Young's slits ("Through which of two slits does a photon 'go' when it contributes to an interference pattern on the other side?"), generalizing that to accommodate the behaviour of an electromagnetic field. What magic that it worked.

Only in the past decade does the physics community seem to have seized on the opportunities marked out by Feynman's triumph. A review of the field last year, by Roland Omnès from Paris-Onze (*Rev. mod. Phys.* **64**, 339; 1992) lists only a sprinkling of references from before the 1980s, and then only for historical reasons. But he makes the point that one of the inspirations of the present interest in what quantum mechanics means were the modern realizations of the Einstein-Rosen-Podolski experiments on non-locality then attempted.

One of the earliest innovations in the field (in 1984) was an attempt by Robert B. Griffiths from the Carnegie-Mellon University in Pittsburgh to generalize Feynman's argument to the notion that, even in quantum mechanics, a mechanical system can have a history. The same Griffiths has now put the point more neatly (*Phys. Rev. Lett.* **70**, 2201; 1993) by asking what may be the equivalent in quantum mechanics of what is called "phase space" in ordinary classical mechanics.

Phase space, of course, is at its simplest merely a space spanned by the position coordinates and the corresponding momenta

of all the particles in a system. The phase space of a single particle moving in one dimension has simply two dimensions, one position and one momentum. A trajectory in phase space is then a line that shows how the momentum of a particle changes with position. A freely moving particle is represented simply by a straight line parallel to the position axis, for example.

Is there a quantum equivalent? Griffiths's construction is a quantum construction from the outset. All systems have eigenstates, represented by the solutions of Schrödinger's equation from which the time-dependent parts have been stripped out. In general, these states are orthogonal to each other; if they are not, they are such that it is always possible to find a finite number of linear combinations of them which are orthogonal to each other and to all others. So there is a kind of phase space, one with an infinite number of dimensions in which a single direction is the equivalent of a single point in classical phase space.

So what happens when such a system, not to begin with in an eigenstate, changes in the course of time? Come what may, it is likely to begin as a combination of a finite number of the eigenstates and so to be represented by a direction enclosed by the vectors that represent its component states. And then it will move with the passage of time, in ways that might be specified.

There is nothing radically new in this. The novelty is that the formulation allows Griffiths to make mathematically explicit the notion that quantum systems can be tackled by considering all possible histories leading to sum specified result, and then adding up their probabilities. It is crucial to his argument that all hypothetical histories must be not merely compatible with the end result, but also such that they do not interfere with each other before they get there.

Nobody will seize upon Griffiths as a way of calculating any but the simplest quantum systems, pairs of photons for example. The interest of his work is that it seems to clarify what many people have been driving at for the past decade, the idea that the probability aspects of quantum mechanics should be built into the foundations of the subject. The implications of this development for the tortuous and sometimes tortured theory of quantum measurement are evidently important.

Even Bohr would have been content with the way the argument has developed.

John Maddox

Certainty from uncertainty

Charles H. Bennett

QUANTUM mechanics is best known for its uncertainty principle, but now physicists and computer scientists have found that quantum mechanics can also serve as a source of certainty. Quantum effects within a computer can be used, in principle, to solve quickly and with complete certainty some kinds of problem that a classical computer could only solve very slowly, or else with a small chance of error¹⁻³.

The extra power of quantum computers arises from the ability of a quantum system, when it is not being observed, to be in many places at the same time. A single photon making its way through a snowball, for example, simultaneously follows all possible optical paths through the snowball, and emerges in a way that is determined by the constructive and destructive interference among all the paths.

Interference

A single quantum computer could similarly, in principle, follow many distinct computation paths all at the same time and produce a final output depending on the interference of all of them. Although a number of obstacles⁴ appear to stand in the way of achieving controllable interference among computation paths in a practical quantum computer, there has been much recent progress in understanding how, in principle, the possibility of such interference makes the theory of quantum computation differ from its somewhat better understood classical counterpart.

An early but probably vain hope was that quantum parallelism might provide a fast way of solving problems such as factoring or the travelling-salesman problem, which appear to be hard in the same way as finding a needle in a haystack is hard, that is because they involve a search for a successful solution among exponentially many candidates. A computer able to test all the candidates in parallel, and signal unambiguously if it found one that worked, would solve these problems exponentially faster than known methods.

It is easy to program a quantum computer to branch out into exponentially many computation paths; the difficult part is getting the paths to interfere in a useful way at the end, so that the answer comes out with a non-negligible probability. This difficulty is illustrated by the factoring problem above. Suppose the quantum computer is programmed to factor a 100-digit number by trying in parallel to divide it by all numbers of fifty digits or fewer. If any of these

approximately 10^{50} computations yields a zero remainder, it will in a sense have solved the problem. But if there is only one successful path, the interference pattern among all the paths, which determines the behaviour of the computer as a whole, will scarcely be affected. Quantum computers cannot amplify an answer found on a single computation to a detectable level because interference is an additive process, to which each path contributes only as much weight as it started out with.

Therefore, a quantum computer's chance of quickly solving a problem of this type is no greater than that of a classical stochastic computer, which can choose a single computation path randomly. For interference to achieve anything useful, it must be applied to problems whose solution depends on many computation paths.

The recent spate of progress in quantum computation stems from a paper by David Deutsch of Oxford University and Richard Jozsa, now at Université de Montréal¹. The authors use a quantum computer to determine global properties of a boolean function f of an n -bit argument, by having the computer branch into 2^n computation paths, one for each value of the argument x . They show that the computer can then cause the paths to interfere in such a way as to determine, quickly and with no chance of error, whether the function f is 'unanimous' ($f(x)=0$ or $f(x)=1$ for all x) or 'balanced' ($f(x)=0$ for exactly half the x values and $f(x)=1$ for the other half).

A classical computer, by contrast, limited to testing arguments one at a time, might find unanimity among all the $f(x)$ it had tested so far, but would still not be able absolutely to rule out the 'balanced' possibility until it had tested more than half the paths, a job requiring an exponentially increasing amount of time. If we are unwilling to tolerate any chance of error, the problem of distinguishing balanced from unanimous functions can be solved quickly on a quantum computer but not on a classical one. On the

other hand, if even a tiny chance of error is allowed, a classical stochastic computer could also solve the problem quickly, by testing a few dozen x values at random, then declaring the function to be unanimous if they all agreed.

In the figure, the top left bar shows a unanimous function and the bar below it a balanced function for a toy example with $n=7$ and 128 x values. Although the global difference between these two bars is obvious to the eye, a classical computer, limited to testing points one at a time, would need to test over half the x values to be sure the upper bar was unanimous. A single quantum computer, by contrast, could quickly and surely detect the global difference by pursuing 128 computation paths in parallel and causing them to interfere.

Ethan Bernstein and Umesh Vazirani at the University of California at Berkeley have strengthened Deutsch and Jozsa's result² by showing that the same quantum computer that distinguishes unanimous from balanced functions can also (and still with no chance of error) distinguish among 2^n-1 kinds of balanced functions. The bars on the right of the figure show two of these functions in the toy example: a quantum computer could distinguish these functions from each other and from 125 other balanced functions, most of which would look confusingly similar to the naked eye. The class of balanced functions distinguishable by quantum computation (the so-called Walsh functions) are of independent mathematical interest, being boolean analogues of the basis functions used in Fourier analysis.

Complexity

What, then, is the relation between the powers of quantum and classical computation? As is well known, classical computational complexity theory suffers from a humiliating inability to answer some of its most basic questions, such as whether the two complexity classes — P, for deterministic polynomial time, and NP, for nondeterministic polynomial time — are equal. In consequence, the factoring and travelling-salesman problems alluded to earlier are merely thought, not known, to be hard. This frustrating situation arises because, in a



A quantum computer can quickly and with certainty distinguish a unanimous function (top, left) from a balanced function (for example, bottom, left). It can also distinguish between a large number of subtly different balanced functions, such as those on the right.

CANCER

Fusion potential for vaccines

Joyce Taylor-Papadimitriou and Frances Balkwill

real problem, alternative computation paths are not independent — for example, in factoring, if a number is divisible by 6 then it will certainly be divisible by 3. Thus one cannot be sure that there is not a much faster way of solving the problem than by blind search: either of these problems might turn out to be more like finding a needle in a department store than finding one in a haystack.

Unable to answer their biggest questions, complexity theorists have retreated to proving two lesser kinds of theorems: oracle results and completeness results. Oracle results concern the power of computers that can ask questions of a black box without being allowed to look inside it. The f functions considered above are an example of an oracle. Because Deutsch and Jozsa's proof does not allow the computer to 'open the box' and examine the instructions for calculating f , it is not an absolute proof that quantum computers are more powerful than classical ones, merely a proof that they are more powerful if certain kinds of f functions exist — functions that are balanced or unanimous, but whose instructions cannot easily be analysed to determine which. Gilles Brassard of Montréal and Andre Berthiaume³, now at Oxford, as well as Bernstein and Vazirani² have proved a number of oracle results based on Deutsch and Jozsa's construction, which characterize in considerable detail the power of oracle-assisted quantum computers relative to their oracle-assisted classical analogues. The bravest of these results is a family of oracle problems which are easy for quantum computers but not for classical ones, even when the latter are allowed to make errors.

The other approach, the completeness approach, eschews oracles and is based instead on finding problems that, although not known to be hard, can be proved to be at least as hard as any other problem in a given class. Thus, in classical complexity theory, the travelling-salesman problem is called NP-complete because all other problems in the class NP can be reduced to it. It may similarly be possible to characterize the power of quantum computation by finding problems that are provably at least as hard as any problem in a given quantum complexity class. □

Charles H. Bennett is at the IBM T. J. Watson Research Centre, Yorktown Heights, New York 10598, USA.

THE development of cancers that express tumour-specific or tumour-associated antigens implies that immune mechanisms have not been effective in rejecting the tumour. Cancer treatments aimed at mobilizing host immunity must therefore not only identify tumour antigens but improve the efficacy of antigen presentation. Recognition of the importance of local cytokine action in increasing antigen presentation has led Tao and Levy (page 755 of this issue¹) to construct a novel and effective vaccine: the vaccine is a fusion protein which combines a well-characterized tumour-specific antigen and a cytokine that stimulates accessory-cell function. The antigen concerned is an antibody corresponding to the specific idiotype expressed on a murine B-cell lymphoma, and the cytokine is murine granulocyte-macrophage colony-stimulating factor (GM-CSF).

The fusion protein has cytokine activity, elicits a strong antibody response, and protects mice from subsequent challenge with tumour cells bearing the specific antigen. Moreover, tumours arising in immunized animals still express the idiotype and there is no evidence of outgrowth of an antigen-negative subclone of the original tumour. Co-administration of the antigen and cytokine, or a fusion protein containing species-specific human GM-CSF, are both ineffective. The fusion of the immunogen with a general carrier protein also generates less effective immunity. These results could have important implications for the design of cancer vaccines, because the administration of a genetically engineered protein offers many advantages over other formulations involving cells or viruses, or both.

How applicable, though, might this approach be to other tumours, particularly human cancers? Obviously, in human clonal B-cell lymphomas, a new vaccine would be required for each patient, but other more widely expressed antigens are currently being developed as immunogens^{2,3}. These will have to be tested as fusion proteins with several cytokines. The generality of the immune-stimulating effect of GM-CSF needs to be examined. This cytokine may not be effective with all antigen-target cell combinations, and evaluation of various cytokines may be necessary to optimize presentation of a particular antigen; a fusion protein of interleukin-2 and a viral antigen, for instance, also seems to be strongly immunogenic⁴. Interestingly, a study of irradiated murine tumour cells, which had been transduced

with retroviruses encoding ten different cytokines or immunomodulators, showed that cells expressing murine GM-CSF were the most effective at stimulating long-lasting and specific immunity⁵.

Another concern is that the fusion protein described by Tao and Levy is tetrameric, comprising two functional GM-CSF molecules, which presumably allows crosslinking of the receptors and generation of a strong functional signal in the effector cells. It is not certain whether the approach would always be as effective using single-chain antigens, or cytokines such as tumour necrosis factor or interferon- γ which are biologically functional as tetramers or dimers.

In the principal human cancers, a similar vaccine could presumably be used as a prophylactic treatment, particularly in patients in remission or those at high risk of developing cancer. There are, however, few data to suggest that vaccination will act against an existing tumour, particularly when the tumour may generate an immunosuppressive microenvironment⁶. Tao and Levy's system would provide a good model for investigating this, and the relative merits of other cytokines as fusion partners.

The use of a tumour vaccine to prevent or delay the onset of malignant disease is one of the ultimate goals of cancer research, but the toxicity of vaccination in a healthy individual is a possible worry, particularly if a cytokine is involved. So it is encouraging to read in the paper¹ that effective immunity was generated with doses as low as 5 nanograms of GM-CSF, well below the doses that stimulate normal or malignant haemopoiesis and are obviously toxic in humans⁷.

The data presented by Tao and Levy open up intriguing possibilities. The applicability of the system to other murine tumour antigens, to human cancer, or indeed to pathogenic organisms, will no doubt now be the subject of a flurry of further papers. □

Joyce Taylor-Papadimitriou and Frances Balkwill are at the Imperial Cancer Research Fund, PO Box 123, London WC2A 3PX, UK.

1. Deutsch, D. & Jozsa, R. *Proc. R. Soc. A* **439**, 553–558 (1992).
2. Bernstein, E. & Vazirani, U. at A. C. M. *Symp. Theory of Computing*, San Diego, 1993 (in the press).
3. Berthiaume, A. & Brassard, G. at *7th IEEE Conf. Structure in Complexity Theory*, 1992 (in the press).
4. Landauer, R. *Phys. Today* 23–29 (May 1991).

1. Tao, M.-H. & Levy, R. *Nature* **362**, 755–758 (1993).
2. van der Bruggen, P. et al. *Science* **254**, 1643–1647 (1991).
3. Jerome, K. R. et al. *Cancer Res.* **51**, 2908–2916 (1991).
4. Hinuma, S. et al. *FEBS Lett.* **288**, 138–142 (1989).
5. Dranoff, G. et al. *Proc. natn. Acad. Sci. U.S.A.* **90**, 3539–3543 (1993).
6. Old, L. J. *Curr. Opin. Immun.* **4**, 603–607 (1992).
7. Lieshke, G. J. & Burgess, A. W. *New Engl. J. Med.* **327**, 28–33 (1992).

Tolerance by exhaustion

Harald von Boehmer

VIRUSES use various strategies to escape protective immune responses and persist in their host. On page 758 of this issue¹, Moskophidis *et al.* show that such persistence can be achieved by brute force: large numbers of viruses produce antigenic loads that force the immune system to give in after a brief and intense struggle, ending with the exhaustion and deletion of cytotoxic T-lymphocytes.

Viruses try to remain in their hosts without killing them too quickly. Viruses such as the lymphocytic choriomeningitis virus (LCMV) — which is the one Moskophidis *et al.* looked at and which is noncytotoxic (that is, does not disrupt or kill cells) — nevertheless induce protective immunity which interferes with the goal: the virus can be eliminated and the associated immunopathology, such as destruction of nervous tissue, may even cause the death of the host. This phenomenon can be observed in animal model systems, and is very infrequently seen in humans exposed to high virus doses. The virus may escape for some time, for instance by infecting cells in the skin that lack cofactors (such as the B7 molecule) required for T-cell stimulation, or through mutation of its immunogenic epitopes. But eventually the immune system may still catch up with it.

It would clearly be most advantageous for the virus if it could induce specific immunological tolerance; the virus would not suffer and the host could still combat other threatening micro-organisms. This goal is achieved by LCMV when it infects mice with an immature immune system during the neonatal period. The viral antigens will delete immature T cells with complementary receptors², much as self-antigens delete immature self-specific T cells. As a result the mice become virus carriers for the rest of their life. Mature T cells are not so easily deleted, but Moskophidis *et al.* show that this can also be achieved.

The authors used high doses of virus, a somewhat artificial approach, but they suggest that a 'proper' virus isolate in a host with the 'proper' major histocompatibility complex (MHC) and non-MHC background will cause specific tolerance even when inoculated in low doses. What is 'proper' in this context remains to be seen. The data presented in the paper show that the 'complete' induction of cytotoxic T cells results initially in a vigorous response followed by their exhaustion and deletion. This can be nicely documented because, as well as normal mice, the authors employed mice with a

transgenic, virus-specific T-cell receptor which allowed them to determine the fate of virus-specific cells.

The induction of specific immunological tolerance in a mature immune system does not only represent an ideal case for viruses, but is also the goal of transplantation surgeons hoping to achieve acceptance of mismatched grafts without having to suppress the entire immune system. It is hardly surprising, therefore, that the results obtained with viruses do not find immunologists totally unprepared. High zone tolerance is an established phenomenon³, and relatively high doses of cell-surface antigens⁴, including transplantation antigens⁵, can induce exhaustion and deletion (and thereby specific tolerance) in a mature immune system of normal mice or mice with transgenic T-cell receptors.

Further experiments should aim at elucidating the molecular mechanisms of exhaustion and deletion of mature T cells. This will help our understanding of whether (or how) viruses establish specific immunological tolerance in adult organisms in real life. It is possible, for instance, that the disappearance of specific cytotoxic T cells during the late stages of AIDS may also be due to exhaustion and deletion. An understanding of this type of tolerance should likewise help immunologists to improve their protocols for inducing tolerance to transplantation antigens. □

Harald von Boehmer is at the Basel Institute for Immunology, 4005 Basel, Switzerland.

1. Moskophidis, D. Lechner, F. Pircher, H. & Zinkernagel, R. M. *Nature* **362**, 758–761 (1993).
2. Pircher, H. P. *et al.* *Nature* **342**, 559–561 (1989).
3. Mitchison, N. A. *Proc. R. Soc.* **161**, 275–293 (1964).
4. Rammensee, H. G., Kroschewski, R. & Frangoulis, B. *Nature* **339**, 541–544 (1989).
5. Rocha, B. & von Boehmer, H. *Science* **251**, 1225–1228 (1991).

COMETS

Great balls of mire

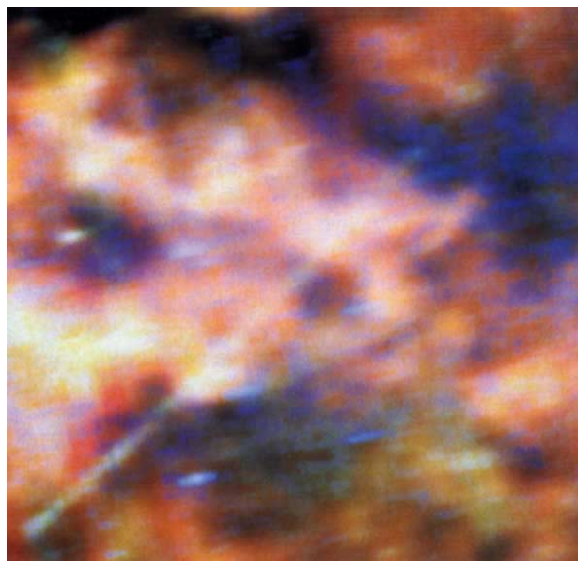
Mark V. Sykes

THE Giotto mission continues to add to its accomplishments since its encounter with comet Halley in 1986. In 1992 the spacecraft flew within 200 km of the nucleus of comet Grigg-Skjellerup and sampled its dust environment. On page 732 of this issue¹, McDonnell and colleagues report that the mass distribution of the dust is dominated by the largest particles, indicating a greater mass loss

rate and dust-to-gas mass ratio than comets were previously thought to have. These results continue to undermine the ice-dominated 'dirty-snowball' model of comets in favour of the 'frozen-mudball' model, dominated by non-volatile, refractory materials.

Dust-to-gas mass ratios corresponding to the dirty-snowball model range between 0.1 and 1. If we were to compress

comet nucleus material so that refractories had a density of 3 g cm⁻³ and volatiles had a density of 1 g cm⁻³, this would give us a nucleus in which 3–33 per cent of the volume consisted of refractory material. Support for this picture comes largely from ground-based observations of dust and at visual wavelengths, sensitive to particles within a factor of ten or so of 1 µm in size. It seems, however, that such observations underestimate the mass fluence of dust from comets, most of which is in the form of larger particles which are difficult to detect at these wavelengths. So much was concluded when Giotto flew by comet Halley², and was strikingly illustrated by



A comet and its dust trail, shown against a background of interstellar dust clouds in this image constructed from scans made by the Infrared Astronomical Satellite.

the discovery by the Infrared Astronomical Satellite (IRAS) of cometary dust trails (see figure)^{3,4}. These latest Giotto observations confirm that mass loss from comet Grigg-Skjellerup follows the same pattern.

Analysis of the IRAS observations⁵ of eight trails indicates that short-period comets lose mass primarily in the form of refractory particles a few millimetres to centimetres in diameter. The average dust-to-gas mass ratio calculated is 3, the upper limit inferred from the Giotto observations of comet Halley⁵ (with a nominal value of 2). Assuming the same densities as above, this corresponds to a comet nucleus that is 75 per cent refractory debris by mass and 50 per cent by volume. A simple backyard experiment demonstrates how apt is the description of such a mixture as a 'mudball'.

This provides us with some insight into the birthplace of short-period comets. Because of their significant ice content, it has long been thought that they must have originated in the outer Solar System or beyond. One suggestion is that some comets were born in molecular clouds⁶. Models of these comets predict a composition dominated by volatile ices with only a small dust component. Dynamical considerations have led investigators⁷ to home in on the regions around the orbits of Uranus and Neptune. Densities determined for both Pluto⁸ and Triton⁹ indicate effective dust-to-gas ratios identical to the average comet values determined from the Infrared Astronomy Satellite and from Giotto. This is exactly what we would expect, as Pluto and Triton would have accumulated from these comets near Neptune's orbit.

If short-period comets are losing mass more rapidly than previously thought, one might expect their lifetime against sublimation to be shorter. A greater fraction of refractory material, however, would allow for the rapid formation of a non-volatile mantle that is difficult to blow off, progressively choking off cometary activity. Such a mantle was apparent in the Giotto images of the Halley nucleus which was near perihelion at the time. When activity had

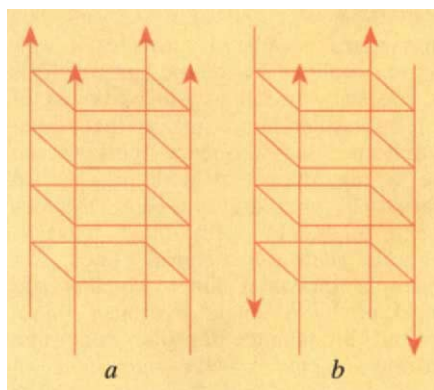
been choked off, the comet would look like an asteroid until sufficient pressure built up from subsurface ices to break through the crust in a burst of resumed activity. The discovery last August¹⁰ that asteroid 4015 was identical to comet Wilson-Harrington, last seen in 1949, provided the first incontrovertible evidence of such 'dormant' comets in the inner Solar System.

RETROVIRUSES

One and one make four

Maxim Frank-Kamenetskii

LIKE other retroviruses, the human immunodeficiency virus (HIV) carries in its virion two identical RNA molecules that are joined together in parallel at their 5' ends. This liaison is believed to be



Parallel (a) and all-antiparallel (b) orientation of four strands (DNA or RNA) forming the quadruplex.

crucial for normal retrovirus function — among other things, the genomic RNA dimer is responsible, at least partly, for the high level of genetic recombination, and consequently for the remarkable genetic variability of retroviruses. It has been demonstrated that, *in vitro*, purified RNAs of HIV-1 and other retroviruses can form dimers^{1,2}. Now come Sundquist and Heaphy, who in a paper in the latest issue of *Proceedings of the National Academy of Sciences*³ present evidence that this parallel dimerization is rooted in a quadruplex structure formed by two identical sites of the RNA molecules. They speculate that the work is the first demonstration that the quadruplex is of biological significance. If it is, then the implications are far-reaching.

DNA quadruplexes have been well characterized by numerous techniques including nuclear magnetic resonance spectroscopy⁴ and X-ray crystallography⁵ (see also my comments on the subject in these columns^{6,7}). Single-stranded telomeric tails may induce dimerization

Giotto's extended mission has undoubtedly added to our knowledge of comet nuclei. The results continue to push us towards a new picture of their composition, and a better understanding of their origin and evolution. □

Mark V. Sykes is at the University of Arizona Steward Observatory, Tucson, Arizona 85721, USA.

of two tailed DNA molecules in the parallel fashion⁸, although all adjacent strands within the quadruplex itself are most probably antiparallel^{5,8,9} (see figure). The DNA quadruplex exhibits remarkable specificity with respect to cations¹⁰; potassium ions specifically stabilize the structure by occupying a highly favourable position within the internal cavity^{5,8,9}. Using NMR data¹¹, it has been shown that RNA oligonucleotides (such as rUG₄U) also form quadruplexes in solution, and that they too are preferentially stabilized by potassium ions.

In their claim that two HIV RNA molecules are joined by a quadruplex structure, Sundquist and Heaphy rely mainly on the observation that the dimers formed exhibit the same cation selectivity as known DNA and RNA quadruplexes; that is, they are preferentially stabilized by potassium ions. Although this argument is not as direct as the usual routes for demonstrating unusual structures in DNA, it seems quite reasonable given that the methods for studying RNA structures lag well behind those available for DNA.

The case would be fully convincing if the region responsible for dimerization carried sequences for which DNA and RNA quadruplexes have been best characterized (such as (T₂G₄) in the case of DNA). Unfortunately, the specific region participating in the quadruplex is yet to be localized, leaving open the possibility that a completely different nucleic acid structure can have the same

- McDonnell, J. A. M. *et al.* *Nature* **362**, 732–734 (1993).
- McDonnell, J. A. M., Lamy, P. & Pankiewicz, G. S. in *Comets in the Post-Halley Era* (eds Newburn, R., Neugebauer, M. & Rahe, J.) 1013–1043 (Kluwer, Dordrecht, 1991).
- Davies, J. K. *et al.* *Nature* **309**, 315–319 (1984).
- Sykes, M. V. *et al.* *Science* **232**, 1115–1117 (1986).
- Sykes, M. V. & Walker, R. G. *Icarus* **95**, 180–210 (1992).
- Greenberg, J. M. in *Comets* (ed. Wilkening, L.) 131–163 (University of Arizona Press, Tucson, 1982).
- Fernandez, J. A. & Ip, W.-H. *Icarus* **47**, 470–479 (1981).
- Tholen, D. *Bull. Am. astr. Soc.* **22**, 1129 (1990).
- Smith, B. A. *et al.* *Science* **246**, 1422–1449 (1989).
- Bowell, E. *IAU Circ. No.* 5585 (1992).

- Darlix, J. L. *et al.* *J. molec. Biol.* **216**, 689–699 (1990).
- Marquet, R. *et al.* *Nucl. Acids Res.* **19**, 2349–2357 (1991).
- Sundquist, W. I. & Heaphy, S. *Proc. natn. Acad. Sci. U.S.A.* **90**, 3393–3397 (1993).
- Smith, F. W. & Feigon, J. *Nature* **356**, 164–168 (1992).
- Kang, C. H. *et al.* *Nature* **356**, 126–131 (1992).
- Frank-Kamenetskii, M. *Nature* **342**, 737 (1989).
- Frank-Kamenetskii, M. *Nature* **356**, 105 (1992).
- Sundquist, W. I. & Klug, A. *Nature* **342**, 825–829 (1989).
- Williamson, J. R., Raghuraman, M. K. & Cech, T. R. *Cell* **59**, 871–880 (1989).
- Pannavaia, T. J. *et al.* *J. Am. chem. Soc.* **100**, 3625–3627 (1974).
- Cheong, C. & Moore, P. B. *Biochemistry* **31**, 8406–8414 (1992).

cation selectivity. But even sceptics will have to admit that, until another structure with such selectivity is found, the hypothesis about quadruplex formation conforms to the principle of Occam's razor.

Sundquist and Heaphy speculate that quadruplex formation may mediate RNA dimerization *in vivo* in HIV and other retroviruses, which would represent the first example of a defined biological function for nucleic acid quadruplexes. *In vitro*, RNA molecules form dimers very slowly, as do DNA quadruplexes, and *in vivo* a protein should facilitate the dimerization. The authors believe that the protein helps to form the

quadruplex, functioning as a kind of RNA chaperone.

If Sundquist and Heaphy are correct that quadruplexes form a structural basis for the 'dangerous liaisons' *in vivo* between two HIV RNA molecules, new possibilities to find effective drugs or vaccines to combat AIDS can be contemplated^{2,3}. For instance, one can imagine that drugs destabilizing or stabilizing quadruplexes would interfere with HIV infection. □

Maxim Frank-Kamenetskii is in the Institute of Molecular Genetics, Russian Academy of Sciences, Moscow 123182, Russia.

INTERMOLECULAR BONDS

The potential to surprise

William Klemperer

HYDROGEN bonding is generally regarded as the strongest and the most directional of the weak intermolecular interactions that cause molecules to form liquids and solids: because of it, water boils at a temperature 160 °C higher than does hydrogen sulphide. Hydrogen bonding is also a primary structure-determining force in biological molecules. Yet results reported on page 735 of this issue¹ show that we know less about this bonding than we should.

Rodham *et al.*¹ have determined the structure of the dimer composed of benzene and ammonia. They find that one proton on the ammonia is directed towards the benzene ring. This discovery is surprising because ammonia, in virtually all other complexes, acts as an exceptionally poor proton donor. In $\text{NH}_3\text{-C}_6\text{H}_6$, although only one proton points towards the benzene plane at any time, the structure is quite dynamic with the ammonia unit undergoing rotation so that all three protons are equivalent. Proton interchange is a general occurrence with complexes of water where monodentate hydrogen-bonded structures are very common.

Advances in spectroscopic techniques over the past two decades have made it possible for chemists to determine the structure of all sorts of molecular complexes. Complexes of virtually any pair of molecules are easily formed by adiabatic expansion, through a pinhole into a vacuum, of a dilute mixture of the pair in an inert gas, usually argon. The effective temperature of the molecular beam formed in such an expansion is below 10 K; thus even the weakest bonds are stable. Spectroscopic measurements of the complexes are made by a variety of techniques. Having the species at very low temperatures, with few quan-

tum states populated, provides a welcome simplification of the spectra. This is especially useful as the weakness of intermolecular bonds means that large-amplitude, low-frequency motions can be a nightmare. In addition, a rich variety of quantum tunnelling motions occur in systems with either identical units or identical chemical bonds. For example, the water dimer, which is held together by a single hydrogen bond, exhibits interchange of proton donor and proton acceptor as well as interchange of bonded proton and free proton of a water unit². The high resolution of pure rotational and far-infrared spectroscopy even allows nuclear hyperfine structure and electric dipole moments to be measured precisely, which is invaluable for learning the orientation of the molecular units with respect to one another³.

A wealth of structures of gas-phase binary complexes is now available, allowing comparisons with first-principles structure calculations and semi-empirical models of intermolecular potentials. It is, of course, an aim of this research to develop an understanding of intermolecular potentials, which will often be applicable to condensed-phase systems as well. For example, rotational spectroscopy⁴ seems to show that the monomer units in the benzene-benzene dimer are arranged in a T-shaped polar structure (not arranged parallel and offset, as some supposed), with a centre-to-centre separation of 4.96 Å. This is almost identical to the geometry of nearest neighbours in liquid and crystalline benzene.

The hydrogen bond is the archetypal stereospecific weak interaction found in complexes. The precise structural characterization of binary hydrogen-bonded complexes allows us to find out exactly

how two units held together by a hydrogen bond orient themselves and how strong the bonds are. Experiments are revealing that hydrogen bonding seems to be subject to many subtle factors. Ammonia forms simple, hydrogen-bonded dimers with the hydrogen halides in the gas phase, but ionic NH_4^+X^- in the crystalline solid phase (X^- being any halide ion). On the other hand, the ammonia homologue trimethylamine forms an ionic complex $(\text{CH}_3)_3\text{NH}^+\text{I}^-$ with hydrogen iodide, even in the gas phase⁵. Of course, in these cases, the ammonia acts as a proton acceptor. In the proton-donating compound studied by Rodham *et al.*¹, $\text{NH}_3\text{-C}_6\text{H}_6$, the hydrogen bond is very weak and relatively floppy, whereas in the trimethylamine-hydrogen-halide complex it is strong and rigid (there is a twentyfold difference in dissociation energy). Another factor now becoming accessible to study is the effect vibration can have on hydrogen bonding⁶. It will be interesting to study the initiation of proton transfer in vibrationally excited hydrogen-bonded complexes.

A further complication is that the linearity of hydrogen bonds appears to depend upon the asymmetry of the molecular unit. For example, the complex OC-HF has a linear equilibrium structure, whereas in OC-HOH , the hydrogen bond deviates by 12° from linear^{7,8}. There is a growing body of evidence that the minimum-energy orientation of the molecules in a complex is a function of the distance of their separation. Indeed, where the orientation oscillates, one finds that there is also a fluctuation in bond length. Even the effects of the unavoidable zero-point vibrations can be accompanied by significant angular reorientation.

The curious, floppy complex studied by Rodham and colleagues touches on many of these issues. It also brings this archetypally chemical-physics field into the domain of biochemistry. As the authors themselves point out, the ammonia-benzene complex is representative of the amino-aromatic interactions that are particularly relevant to protein structure. □

William Klemperer is in the Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138, USA.

1. Rodham, D. A. *et al.* *Nature* **362**, 735–737 (1993).
2. Fraser, G. F. *Int. Rev. Phys. Chem.* **10**, 189–206 (1991).
3. Saykally, R. J. *Acc. Chem. Res.* **22**, 295–300 (1989).
4. Arunan, E. & Gutowsky, H. S. *J. Chem. Phys.* **98**, 4294–4296 (1993).
5. Legon, A. & Arunan, E. *J. Chem. Phys.* (in the press).
6. Suhm, M. A., Farrell, J. T., McIlroy, A. & Nesbitt, D. J. *J. Chem. Phys.* **97**, 5341–5354 (1992).
7. Yaron, D. *et al.* *J. Chem. Phys.* **92**, 7095–7109 (1990).
8. Bumgarner, R. E., Suzuki, S., Stockman, P. A., Green, P. G. & Blake, G. A. *Chem. Phys. Lett.* **176**, 123–127 (1991).

Theory fits the bill in the Galápagos Islands

Jeremy J. D. Greenwood

THOSE who criticize the theory of evolution by natural selection often do so on the grounds that it is impossible to test it by making quantitative predictions. Rosemary and Peter Grant now show this view to be wrong¹. Not only have they observed an environmental change and its effects on the feeding ecology of their study species but they have measured the resultant selective forces and, using previous estimates of the heritability of the traits involved, have predicted the resultant evolutionary changes. Their predictions have been precisely correct.

The Grants' study species was *Geospiza fortis*, one of the ground finches of the Galápagos Islands; this is a family that has contributed much to the study of evolutionary ecology through David Lack's famous book² and the subsequent work of the Grants and others. The population on Daphne Major, an undisturbed island of only 0.34 km² in extent, has been studied closely since 1976, as has the abundance of the seeds on which the birds depend and other aspects of the island's ecology. During 1982–83 there was a severe El Niño event, perhaps the greatest oceanographic and climate disturbance in the Pacific for a hundred years. It brought eight months of heavy rain, totalling over 1,300 mm (the usual annual total is less than one-sixth of this). Not surprisingly, there

were profound effects on the vegetation. In particular, the abundance of large, hard seeds declined substantially but that of small, soft seeds increased tenfold, so that the total biomass of seeds available was much greater. The changes in relative abundance of large and small seeds persisted until at least 1991.

The response of *G. fortis* to the increased food supply was dramatic. The 1982–83 rains gave rise to prolonged breeding. Indeed, some birds that were hatched early in the event were themselves breeding before the end of it. Numbers increased rapidly. Afterwards, with a return to normal, dry conditions, there was no further breeding until 1987. Birds died in the intervening four years, so that only 37 per cent of those alive in 1983 bred in 1987. Furthermore, that 37 per cent did not constitute a random sample — they had longer and narrower bills than the 1983 average. This fits in with observations on feeding behaviour in this species: when small seeds are more abundant, they make up a greater proportion of the diet of the birds (especially of individuals with narrow bills). As expected, there was a marked shift to small seeds during and after the El Niño event.

That the 1982–83 event resulted in marked selection on the bill size of *G. fortis* has been known for some time³.

The Grants have now updated their measures of selection, predicted its outcome, and tested their prediction. The magnitude of selection on bill width and bill length was measured by comparing the mean values for birds that survived to breed in 1987 with those of the 1984 population. In earlier studies, Boag⁴ had measured the heritability of bill width and bill length in this population, as well as the genetic covariance between these characteristics (wide bills tend also to be long). Using standard quantitative genetics theory, it was therefore possible to predict the outcome of the observed selection; that is, the difference in means between the original population and the offspring of those that bred. For bill width, a decrease of 0.13

mm was predicted. In the event, there was a highly significant decline of 0.12 mm — a remarkably close fit.

The prediction for bill length was that it should increase, but only by 0.01 mm as the direct selection for long bills was almost balanced by selection for narrow bills (which tend to be short). In the event, there was a non-significant increase of 0.06 mm in mean bill length.

These results are supported by those for the second species on the island, *G.*



scandens. This feeds on cactus seeds, which declined markedly in abundance after 1983. As a result, there was a substantial decrease in *G. scandens* numbers because this species did not shift its diet; but, precisely because the diet was unchanged, the decline was not selective. No change in bill dimensions was therefore predicted for this species, and none occurred.

As well as contributing to our understanding of evolution, this study has object lessons for the way we do ecology. For too long, even ecology has been dominated by the short-term nature of science funding and by the need, if one is to make a mark in a competitive world, to produce results quickly. Fortunately, the value of long-term studies is now almost universally recognized, thanks to those such as the Grants who have kept such studies going against the odds and who, in doing so, have shown how essential is a long-term approach if we are to have a true understanding of the natural world. □

Jeremy J. D. Greenwood is at the British Trust for Ornithology, The Nunnery, Thetford, Norfolk IP24 2PU, UK.



The finch connection — *Geospiza scandens*, with both female and male, and (above right) a lone *G. fortis*. The illustrations, by John Gould, first appeared in the *Zoology of the Voyage of the H.M.S. 'Beagle' (1841)*, and are reproduced from the 1983 reissue of ref. 2.

1. Grant, B. R. & Grant, P. R. *Proc. R. Soc.* **B251**, 111–117 (1993).
2. Lack, D. *Darwin's Finches: An Essay on the General Biological Theory of Evolution* (Cambridge University Press, 1947).
3. Gibbs, H. L. & Grant, P. R. *Nature* **327**, 511–513 (1987).
4. Boag, P. T. *Evolution* **37**, 877–894 (1983).

Forming galaxies spotted?

Richard Ellis

FINDING one primordial galaxy would be a cause for congratulations, but finding a group of them would be grounds for celebration. Alan Dressler and collaborators claim to have done just that, using the unique imaging capabilities of the Hubble Space Telescope (*Astrophys. J.* **404**, L45–L49; 1993). The putative group was found as the astronomers examined a foreground galaxy cluster which was the original target of their study. This underlines the astronomer's watchword — to expect the unexpected.

One of the main questions in cosmology is when the first galaxies formed and how they subsequently evolved to their present diverse forms. To address this question, large ground-based telescopes are used as time machines. The finite velocity of light enables astronomers to probe ancient cosmic history by looking to enormous depths. Previous searches for young or 'primaeval' galaxies have been hampered by a lack of consensus as to the period during which such objects would be seen (that is the epoch of

formation) as well as some uncertainty as to what they would look like. Naively one might expect a collapsing proto-galactic cloud to be an extended region displaying intense emission lines of hydrogen heated by the young forming stars. Although some promising examples have been identified from radio surveys, their rarity indicates that these are unlikely to be the precursors of the bulk of the more numerous galaxies we see today.

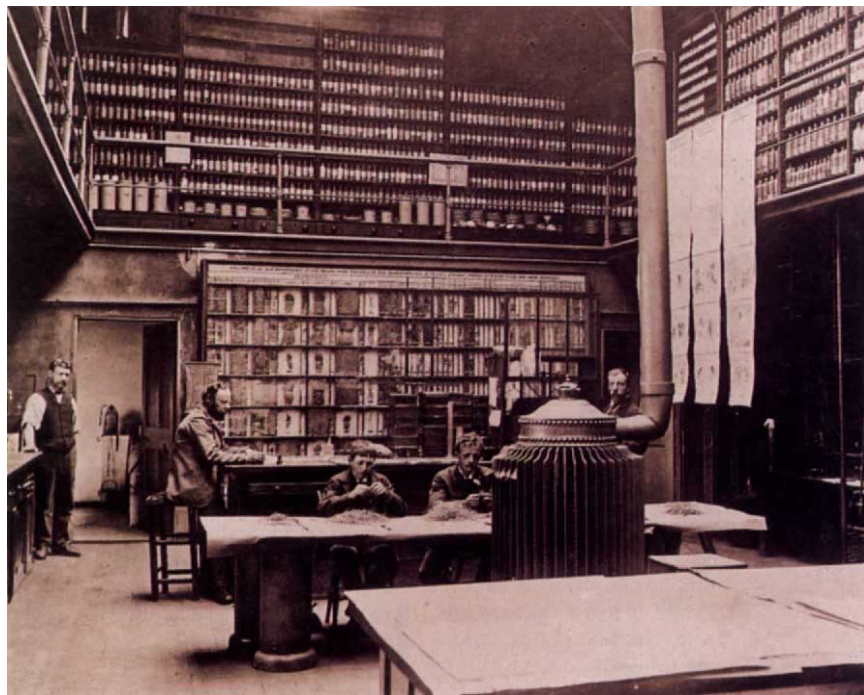
The Hubble Telescope was not expected to play a significant part in this field, but Dressler's team argues that this reasoning prejudices what primaeval galaxies may look like. Suppose that, during the initial formation phase, the active sites of star formation are distributed as points of light within a dim cloud. The Hubble Telescope, even with its current optical difficulties, would be well-placed to spot these points of light and distinguish primaeval galaxies from the more numerous extended images of the foreground population.

The evidence presented for Dressler

and co-workers' "cluster of nascent galaxies" is intriguing but more data will certainly be needed before sceptics are convinced of its importance. Within the image of the foreground cluster they originally set out to study, a small grouping of about 30 faint sources is seen whose colours are quite blue. A significant feature of these sources, which distinguishes them from the galaxies in the main cluster, is their small apparent size and bright point-like nuclei. Such detail is difficult to secure with even the best ground-based telescopes.

Tantalizingly, the distance to one of these sources is already known. By chance, during a spectroscopic survey with the Palomar 200-inch telescope, Dressler and James Gunn found one to be a quasar about four times farther away than the foreground cluster. The quasar lies near the centre of the distribution of faint sources and it is tempting to imagine that the group forms a physical association. If correct, the blue sources would be 11 billion light years distant, corresponding to a 'look-back time' about 80 per cent of that to the Big Bang itself. The group would be comparable in diameter to that of the Local Group containing the Milky Way and

Eminent Victorians and science at the grass roots



Rothamsted at the turn of the century: grass samples being sorted.

GILBERT and Sullivan are perhaps the best known of great Victorian duos, but Gilbert and Lawes also have a strong claim on history. On 1 June 1843 John Lawes employed Henry Gilbert, a chemist, to help with various experiments on his estate at Rothamsted, near Harpenden,

just north of London. The event marks the advent of scientific rigour in agriculture — the two undertook a vast range of investigations into soil chemistry, plant and animal nutrition and crop production, in a partnership that lasted for well over half a century.

This summer Rothamsted Experimental Station, the modern incarnation of Lawes and Gilbert's research enterprise, puts out the bunting and celebrates 150 years since its founders got together. There is quite a bit to celebrate. Among the medals on Rothamsted's chest are that it was the home for the genesis of R. A. Fisher's revolution in statistical practice; for F. C. Bawden and N. W. Pirie, whose studies of tobacco mosaic virus were one of the strands woven into the field now known as molecular biology; and for the discovery of the pyrethroid insecticides.

In accord with the legacy of Lawes and Gilbert, one of the main characteristics of research at Rothamsted has been the acquisition of long-term, continuous data sets (that will be the theme of the 150th anniversary conference to be held on 12–14 July). The oldest come from the Broadbalk experiment, in which winter wheat has been sown and harvested continuously on the same experimental plots since 1843.

Rothamsted is supported in large part through the Agricultural and Food Research Council, and now also pulls in a healthy slice of contract work. Its continued existence seems assured, but inevitably money is another theme of the anniversary year — the aim is to raise £7 million for Rothamsted International, which will provide fellowships for scientists from developing countries. T. L.

the Andromeda spiral. In the 'bottom up' picture for the formation of cosmic structures, such groups are expected to be the building blocks from which large rich clusters are made.

The assignment of a large cosmic distance to the sources depends critically on the question of their association with the quasar. Were the 30 images the only ones of their kind in the area, this might be convincing. However, a high proportion of the faintest images in the field are blue, including many nowhere near the quasar. The authors state that these appear to be "visually indistinguishable" from those clustered around the quasar. As the sources close to the quasar are not unique in their morphological nature, the claim for an association is considerably weakened. However, if most of the faint blue sources in the field were distant, the high surface density of unusual sources might then justify the label *primaevial galaxy*.

The other ingredient to Dressler and co-workers' discovery is the unusual point-like nature of the faint sources. By virtue of its high spatial resolution, the Hubble Telescope is quite likely to present images of an unfamiliar nature. Blue colours indicate systems with active star formation over a wide range in distance. Possibly many fields of view will include knot-like regions of star-formation when viewed with Hubble. Such distinguishing features do not necessarily imply sources at enormous distances, which is a pre-condition for the term 'nascent galaxies'.

The critical step forward in resolving the ambiguities must be to verify the physical association of the sources close to the quasar by measuring their redshifts spectroscopically. Distances should also be measured for some of the other blue sources in the field. For the faintest examples, this will be a demanding observational programme even with the new generation of 8–10-metre telescopes coming on line. However, 4-metre telescopes have already surveyed blue sources in other areas of the sky representative of the brightest examples in the Hubble image. Significantly, these surveys found most of the bluest galaxies are low-luminosity examples at modest distances. Many of them are compact and could be the population identified by Dressler and co-workers. The apparent association with the quasar would then have to be false. So the onus is on Dressler and his colleagues to complete the ground-based follow-up work. Only then will their claim for one of the holy grails of observational cosmology be justified. □

Richard Ellis is in the Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK.

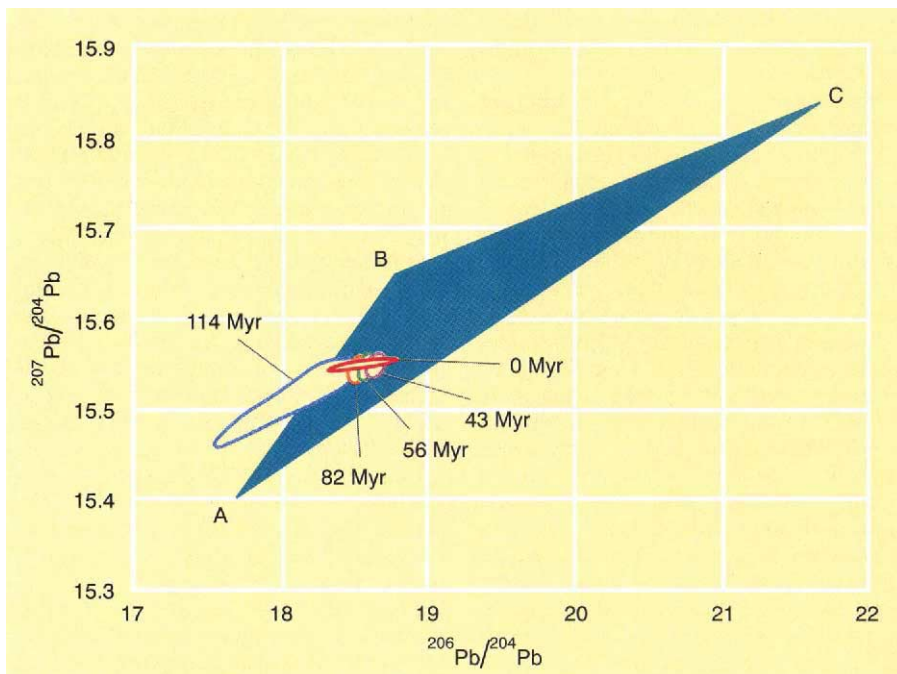
How the Earth's mantle could lie about its age

Richard W. Carlson

RADIOACTIVE decay has left its mark on the isotopic features of the Earth's rocks. With certain assumptions, these features can be used to infer the age of geochemical events that fractionated the parent (radioactive) from daughter (stable) isotope. Application of these

of the Earth's interior. How well is the mantle stirred by convection? Is the mantle compositionally layered with convection restricted within layers? What is the fate of crustal material reinjected into the mantle by plate subduction?

One approach for understanding



Lead-isotope variations in oceanic basalts mostly lie within the triangle A–B–C. Because ^{207}Pb is the product of relatively short-lived ^{235}U , large vertical variations on the graph correspond to ancient changes in U/Pb in the sources of these basalts. Large horizontal variations may be the result of more recent processes if U is suitably enriched over Pb. The elliptical fields encompass basalts from the Ninetyeast Ridge grouped according to their age. The nearly horizontal progression of data is the basis of the young age estimated by Class *et al.* for the formation of the plume source supplying this volcanic ridge.

techniques to the geochemical history of the mantle suggests the presence of geochemically distinct materials that have remained isolated from one another for more than 2.5 billion years. But on page 715 of this issue, C. Class, S. L. Goldstein, S. J. G. Galer and D. Weis take a new look at this information, and suggest instead that the mantle heterogeneity is not as old as it looks.

Understanding the geochemical evolution of the Earth is a path to defining the history of the planet, for it provides information on the mechanism and timing of the Earth's separation into its compositionally distinct envelopes of core, mantle, crust, hydrosphere and atmosphere. Further, the identification of geochemical heterogeneity in the mantle, particularly defining the longevity of that heterogeneity, provides clues to the thermal and mechanical behaviour

compositional variation in the Earth's interior is through the study of melts of the mantle erupted onto its surface. Prime sites for these studies are the world's ocean-ridge systems and the oceanic intraplate volcanoes, such as Hawaii or the Azores. Because they erupt through young oceanic crust, both ridge and ocean-island lavas are immune to overprinting by the chemically distinct and often ancient materials that one finds in continental crust and, hence, carry the unmodified isotopic signature of their source in the mantle.

Isotopic variations observed in oceanic basalts show that there are materials with distinct geochemical histories in the mantle. For example, in plots of the ratios $^{207}\text{Pb}/^{204}\text{Pb}$ against $^{206}\text{Pb}/^{204}\text{Pb}$ (see figure; ^{207}Pb and ^{206}Pb are the decay products of radioactive ^{235}U and ^{238}U , respectively), oceanic basalts define a

roughly triangular field. One interpretation of this field is that the apices of the triangle are three 'end-member' geochemical components in the mantle. Assigning a mechanism of origin to these end-members is a subject of much debate. Point A is the source composition of most ocean-ridge basalts, and is depleted in those elements that preferentially have been transferred into the continental crust. Points B and C are often attributed to the influence, respectively, of old continental crust and basaltic oceanic crust injected back into the mantle by plate subduction. Mixing between these end-members then produces erupted lavas with isotopic compositions within the triangular field.

With some assumptions, the observed isotopic variation can be used to infer the relative ages of the end-member components. As the decay half-life of ^{235}U is less than one-sixth of the age of the Earth, most of the ^{235}U originally present has decayed away. Consequently, large variations in $^{207}\text{Pb}/^{204}\text{Pb}$ (giving steep slopes in the diagram) correspond to ancient variations in U/Pb whereas recent fractionation of U from Pb will be incapable of producing large changes in $^{207}\text{Pb}/^{204}\text{Pb}$. Thus, the isotopic distinction between ocean-ridge basalts, which generally plot close to point A on the figure, and those ocean-island basalts plotting near point B reflects source regions that have maintained distinct U/Pb ratios for over 2.5 billion years. Neither source region, however, is required to have been present in the mantle for this long. For example, sediments eroded from continents will carry with them the old mean age of the continents even though their reintroduction to the mantle may have occurred fairly recently through plate subduction.

At least two different mechanisms, with distinct consequences for the age of mantle geochemical heterogeneity, can explain the origin of the basalts that appear near point C in the figure. If the source of C-type lavas was derived from the source of ocean-ridge basalts (point A), the slope of the line connecting A and C corresponds to a fractionation age of 1.7 billion years. On the other hand, if the C-type source were generated from sources like those at point B, then significantly less time is required to generate the observed Pb isotope characteristics. If done in a short time, however, the large increase in $^{206}\text{Pb}/^{204}\text{Pb}$ from B to C would indicate a dramatic increase in U/Pb ratio in the C-type source, so dramatic that most workers favour the 1.7-billion-year interpretation for the age of source C.

The study of Class *et al.* provides a new approach for deciding between these alternatives. In the ocean basins are several long volcanic chains believed

to be formed as crustal plates move over stationary plumes of hot material rising from the deep mantle. Class *et al.* reasoned that if the composition of the plume supplying the volcanism associated with one chain remained constant over the 100 million years or more that volcanism persists, examination of the isotopic changes along the chain, in basalts of widely varying age, could directly provide the parent-daughter ratios in the plume.

Their study of Pb isotope data for the basalts of the Ninetyeast Ridge, which connects the 117-million-year-old basalts of eastern India to the recent volcanism of Heard Island in the southern Indian Ocean, shows that $^{206}\text{Pb}/^{204}\text{Pb}$ increases significantly in going from old to young basalts along the chain. With the exception of the oldest basalts, $^{207}\text{Pb}/^{204}\text{Pb}$ is nearly invariant in the samples from this ridge (see figure). The conventional explanation for such a data array is that it represents mixing between a point on the A-B join with one on the A-C join. Hence, the line would have no age significance. Class *et al.*, however, argue that this array represents radiogenic increase of ^{206}Pb in the plume source over the 114-million-year history recorded by these basalts. To create the observed variation, Class *et al.* calculate that a U/Pb ratio of about 0.5 is required in the plume. This is indeed large compared with the values of about 0.1 typical of the ocean-ridge basalt source.

From this observation, Class *et al.* conclude that the geochemical fractionation responsible for producing the high $^{206}\text{Pb}/^{204}\text{Pb}$ component in the Ninetyeast Ridge plume happened not more than a few hundred million years ago. The process most likely to create high U/Pb in the plume source is partial melting near the Earth's surface. A common model for mantle plumes, however, has them forming as hot 'balloons' from a thermal boundary layer at the core-mantle boundary. If Class *et al.* are right, the young inferred age of the Ninetyeast Ridge plume source requires either a shallow plume source, or an unexpectedly rapid route for surface material to get to the core-mantle boundary and back again. A young plume age also contradicts the now long-held assumption that geochemical heterogeneity in the convecting mantle can be preserved for billions of years. Instead, the argument suggests a well-stirred mantle that takes on the disguise of old age by subduction of the ancient materials of the cold, convectively stagnant continental crust and upper mantle. □

Richard W. Carlson is in the Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road NW, Washington, DC 20015, USA.

Textile topology

TEXTILE fabrics, says Daedalus, are one of the cleverest of inventions. The fibres are held together only by friction, yet the structure keeps its integrity. Even a knitted garment, which in principle consists of a single fibre knotted into an intricate two-dimensional manifold, holds together well despite its flexibility. Its one disadvantage is that it is hard to repair. Once a hole has worn in it, it must be crudely patched or discarded.

Daedalus now plans to get round this problem. Imagine, he says, a knitted garment made of one single fibre; and imagine that the knitting process returns to its origin so that the two ends of the fibre are adjacent. Fuse the ends together with a smooth, knotless join. The garment is now knitted from a single endless loop. This loop can then circulate round the garment, simply by traversing along its own length. The garment will retain its shape. But an area of high wear will be steadily renewed by the arrival of new fibre; the worn fibre will move away and be dispersed into less heavily abraded areas. Instead of developing a hole, the garment will wear evenly over its entire surface. After an amazingly long life, it will finally fall to pieces.

DREADCO fibre technologists are now elaborating this notion. They are devising a fibre with many tiny barbs, so that it can move easily in one direction, but cannot go back. The random slippage at any fibre junction now becomes directional. Daedalus hopes that the natural movements of the wearer will induce the circular fibre to travel round its loop, slowly but steadily. But to speed things up, he is inventing a special cleaning fluid with exactly the same surface-energy as the fibre. This eliminates friction from its many crossing-points. By submerging the garment in this fluid and subjecting it to ultrasonic vibro-cleaning, the loop is made to circulate at high speed, distributing and dispersing the accumulated wear.

DREADCO's knitted-loop garments should sell like hot cakes. The economical will value their sheer endurance. The fashionable will love their changing appearance. For the pattern printed on the fabric will slowly shimmer and disperse as the loop winds around, displacing its various elements. At unpredictable intervals it will reappear, upside-down, backwards, or in still stranger topological transformations. At every wearing, a knitted-loop garment will renew its novelty.

David Jones

Magnetic compass orientation

SIR — Recent behavioural evidence suggests that the mechanism underlying magnetic compass orientation in newts is light-dependent^{1,2}. Among these results are: (1) when newts are orienting magnetically, the direction in which they are moving changes when the animals are exposed to certain specific wavelengths of light¹; and (2) newts orient magnetically under full-spectrum light but not in the absence of visible light². Such results are consistent with models proposing that magnetoreception involves a modulation of the response of retinal photoreceptors to light, and therefore cannot occur in darkness.

Light is not necessary, however, for magnetic compass orientation in all vertebrates. Recent experiments with loggerhead sea turtle hatchlings have demonstrated that tethered turtles swimming in complete darkness can orient to the Earth's magnetic field³. I have recently repeated this finding in both loggerhead hatchlings and in a second species of marine turtle, the leatherback *Dermochelys coriacea* (unpublished data). Representatives of two other vertebrate classes (fish⁴ and mammals⁵) also probably orient magnetically in darkness, as do several invertebrates^{6,7}. These results suggest that light-independent magnetic compasses are phylogenetically widespread.

The light-independent compass of sea turtles and the light-dependent compass of newts may or may not rely on the same transduction process. The turtle compass⁸, however, resembles the magnetic compass of shorewards-orienting newts⁹ in that both compasses are axial and based on field-line inclination rather than on field polarity. Thus, light-dependence is apparently not a universal feature of vertebrate magnetic compasses, of invertebrate magnetic compasses, or of inclination compasses.

Magnetoreception in darkness may still occur in photoreceptors if the transduction process relies on magnetic field-dependent biochemical reactions¹⁰ that are independent of light. In principle, such reactions could occur anywhere in

the body, but the photoreceptors are an appealing location because the retina provides an ordered array of receptor molecules. At present, however, there is no direct evidence for such a light-independent biochemical transduction mechanism, either in photoreceptors or elsewhere. Further research is required to determine whether magnetic compasses capable of functioning in darkness rely on different receptors from the apparently light-dependent magnetic compass of newts, or whether all of these compasses in fact share similar underlying receptor mechanisms.

Kenneth J. Lohmann

Department of Biology,
University of North Carolina,
Chapel Hill,
North Carolina 27599-3280, USA

Electrochemical mercury detection

SIR — Concern about the environmental dangers of mercury pollution has led to remarkable efforts towards developing analytical methods for this metal. To determine the environmental levels of mercury, which can be of the order of a few parts per trillion (10^{12}) (p.p.t.), highly sensitive and selective methods must be used. However, most approaches require a preconcentration or a pretreatment step owing to their low sensitivity or selectivity^{1,2}. We report here on the application of a glassy carbon electrode (GCE) spin-coated with 4,7,13,16,21,24-hexaoxa-1,10-diazabicyclo[8.8.8]hexacosane (Kryptofix-222) for very selective determination of ultra low levels ($<10^{-12}$ M) of mercury.

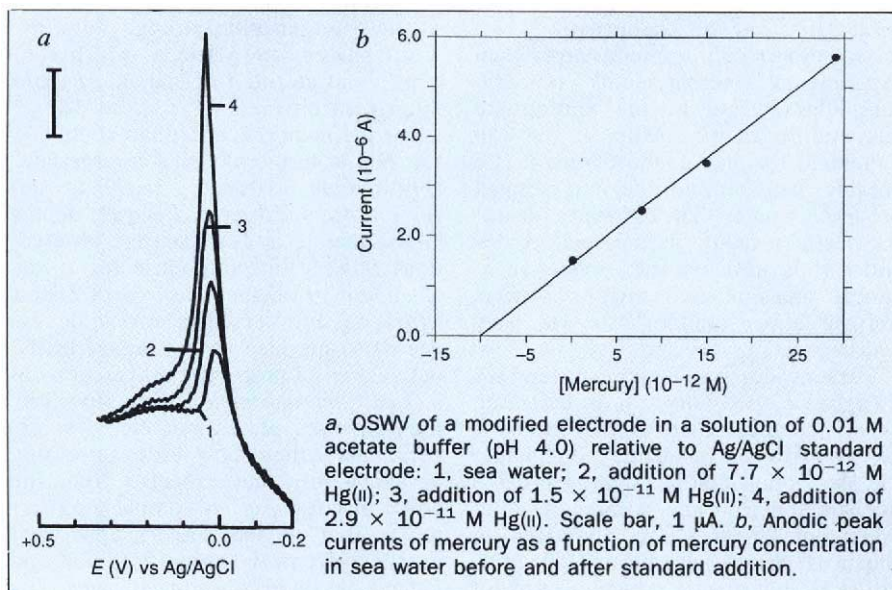
The figure shows the voltammetric analysis of sea water (Mediterranean Sea, Israel) obtained with a modified electrode without further treatment. The anodic stripping Osteryoung square wave voltammograms (OSWV) reveal a clear anodic peak at 0.1 V, associated with the oxidation of predeposited mercury (5 min deposition time at -0.5 V). Electrodes were electrochemically regenerated after each experiment (soaking in solution for 1 min at 0.3 V) and, as a result, could be applied for numerous experiments. The concentration of mercury in sea water was estimated as $(1.01 \pm 0.03) \times 10^{-11}$ M (2 p.p.t.) using a standard addition method.

Optimizing the various parameters that affect the sensitivity of mercury determination, such as potential and time of deposition, resulted in a linear dependence between the stripping peak current and concentration of Hg(II) in the range of 1.5×10^{-12} to 1.2×10^{-11} M. The detection limit is less than 10^{-12} M (0.2 p.p.t.) with a relative standard deviation of 3.3%.

To confirm our results as well as to verify the application of the developed method to the analysis of natural waters, simultaneous analyses have been accomplished by our method and by cold vapour flameless atomic absorption. Excellent agreement between the results obtained from both methods has been found. For example, the concentrations of mercury in two samples were 0.75 ± 0.06 and 82.2 ± 0.1 p.p.b. as compared with 0.90 and 89.5 p.p.b., respectively, obtained by flameless atomic absorption. These samples had to be diluted by 450 and 45,000 times, respectively, before voltammetric analysis could be pursued.

The fact that the results obtained from the analysis of natural samples by our method were in good agreement with another analytical method suggests that

1. Phillips, J. B. & Borland, S. C. *Nature* **359**, 142–144 (1992).
2. Phillips, J. B. & Borland, S. C. *Anim. Behav.* **44**, 796–797 (1992).
3. Lohmann, K. J. *J. exp. Biol.* **155**, 37–49 (1991).
4. Quinn, T. P. *J. comp. Physiol.* **A137**, 243–248 (1980).
5. Burda, H., Marhold, S., Westenberger, T., Wiltshko, R. & Wiltshko, W. *Experientia* **46**, 528–530 (1990).
6. Arendse, M. C. *Nature* **274**, 358–362 (1978).
7. Lohmann, K. J. & Willows, A. O. D. *Science* **235**, 331–334 (1987).
8. Light, P., Salmon, M. & Lohmann, K. J. *J. exp. Biol.* (in the press).
9. Phillips, J. B. *Science* **233**, 765–767 (1986).
10. Schulten, K. & Windemuth, A. in *Biophysical Effects of Steady Magnetic Fields* (eds Maret, G., Boccard, N. & Klepner, J.) 99–106 (Springer, Berlin, 1986).



a, OSWV of a modified electrode in a solution of 0.01 M acetate buffer (pH 4.0) relative to Ag/AgCl standard electrode: 1, sea water; 2, addition of 7.7×10^{-12} M Hg(II); 3, addition of 1.5×10^{-11} M Hg(II); 4, addition of 2.9×10^{-11} M Hg(II). Scale bar, 1 μ A. b, Anodic peak currents of mercury as a function of mercury concentration in sea water before and after standard addition.

our reported approach is also highly selective. Indeed, the determination of 10^{-11} M of Hg(II) ions was not affected by the addition of 10^{-5} M zinc, copper, lead or manganese or 10^{-7} M silver. Furthermore, the introduction of 0.1 M acetate, chloride, bromide or thiocyanate anions did not interfere with mercury analysis.

Such modified electrodes offer a simple, fast and reliable method for monitoring mercury in our environment. To the best of our knowledge, the developed method represents the lowest detection limit ever reported for Hg(II) using an electrochemical technique, which can be compared to the detection limit of expensive and relatively complicated methods³ such as inductively coupled plasma/mass spectrometry.

Iva Turyan

Daniel Mandler

Department of Inorganic and Analytical Chemistry,

The Hebrew University of Jerusalem, Jerusalem 91904, Israel

1. Bloom, N. S. & Crecelius, E. A. *Mar. Chem.* **14**, 49–59 (1983).
2. Szakacs, O., László, A. & Horváth, Z. *Analytica chim. Acta* **121**, 219–224 (1980).
3. Haraldsson, C., Westerlund, S. & Öhman, P. *Analytica chim. Acta* **221**, 77–84 (1989).

Conduction by mantle hydrogen

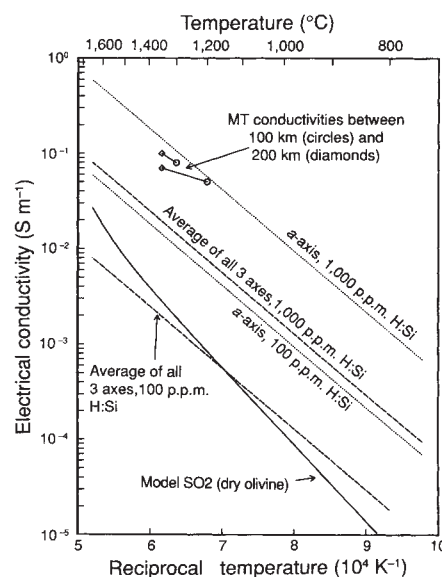
SIR — Karato's suggestion¹ that hydrogen dissolved in olivine might contribute substantially to the electrical conductivity of the upper mantle has gained a sympathetic ear in the electromagnetic community. It has long been believed that the upper mantle at a depth of around 200 km has a relatively high electrical conductivity of about 0.1 S m^{-1} , based to a large extent on inversions by Oldenburg of seafloor electrical data², but this conductivity is one to two orders of magnitude larger than experimental measurements of dry, subsolidus olivine in the appropriate temperature range³. Although the requirement for high conductivities in the oceanic mantle has now been questioned on the grounds that coastlines distort the seafloor data⁴, partial melt, water and carbon have all been suggested at various times as conductivity enhancers for the upper mantle; now we have hydrogen.

Karato's idea, while excellent food for thought, is speculative. It is based on experimental evidence for large diffusivities of hydrogen in olivine⁵, but makes the assumption that the hydrogen exists as charged defects and extrapolates solubilities of 20–80 p.p.m. hydrogen to silicon (H:Si) at 300 MPa to 100–1,000 p.p.m. H:Si at mantle pressures of about

4 GPa (130 km depth). While observations of hydrogen content in olivines from mantle xenoliths are only between 0 and 80 p.p.m.⁶, the large diffusivities of hydrogen imply that any higher concentrations of hydrogen would be lost during eruption.

Crystallographic anisotropy of diffusivity has been largely ignored by proponents of the hydrogen conduction hypothesis. Mackwell and Kohlstedt⁵ report diffusivities for all three crystallographic axes, with $D_a \approx 10D_c \approx 100D_b$. Karato used only the *a*-axis diffusivities in his calculations, and while he acknowledged that the effect of anisotropy would reduce the conductivity of an isotropic mantle, this was not quantified. I have applied various mixing laws⁷ to predict the electrical conductivity of an isotropic mixture of all three axes, computing the parallel and series bounds of the conductivities along the three axes as implied by Karato's work, as well as the much more realistic geometric mean and effective medium average (an average of the series and parallel bounds, and shown in the figure). Also shown in the figure are results for the SO₂ model of isotropic mantle conductivity based on laboratory measurements on olivine³. One sees that for 100 p.p.m. hydrogen, the average conductivity of an isotropic mantle would be dominated by the conventional model of conductivity above 1,000 °C; 1,000 p.p.m. hydrogen produces conductivities which are only 5–6 times greater than the 'dry' model at 1,300 °C. To achieve inferred mantle conductivities using the hydrogen model, one needs about 8,000–10,000 p.p.m. H:Si. This might not be impossible; at 3 GPa hydrogen solubility in olivine could be as high as 5,000 p.p.m.⁸, based on extrapolation of the observed dependence of hydrogen solubility on water fugacity, hydrogen fugacity and oxygen fugacity at 50–300 MPa. However, water is likely to partition strongly into hydrous phases, or pyroxene, which contains about an order of magnitude more water than olivine⁶.

For the mantle deeper than about 200 km, the isotropic model is appropriate. Deformation texture is likely to develop to some extent at shallower depths in the mantle and, as seismic observations suggest a preference for *a*-axis alignment in olivine, the axis of fastest hydrogen diffusion, this would be expected to produce an anisotropy in the bulk electrical properties if hydrogen did indeed dominate conduction. However, the responses of seafloor electrical experiments either show little anisotropy or show only that expected from the effect of coastlines. A complicating factor is that in the cooler, uppermost mantle electrical properties will no longer be dominated by silicates, be-



cause the activation energy for conduction by hydrogen (1.3 eV) is similar to that in dry olivine (1.6 eV), which has an electrical conductivity less than 10^{-5} S m^{-1} below 700 °C. Although electrical conduction in mantle olivine dominated by hydrogen has not been disproved, the large concentrations required to enhance bulk conductivity sufficiently, a probable low partition coefficient for water in olivine, and electrical models not requiring high mantle conductivities, conspire to make it less likely than previously supposed.

Steven Constable
Institute of Geophysics and Planetary Physics,
University of California, San Diego,
La Jolla, California 92093-0225, USA

cause the activation energy for conduction by hydrogen (1.3 eV) is similar to that in dry olivine (1.6 eV), which has an electrical conductivity less than 10^{-5} S m^{-1} below 700 °C. Although electrical conduction in mantle olivine dominated by hydrogen has not been disproved, the large concentrations required to enhance bulk conductivity sufficiently, a probable low partition coefficient for water in olivine, and electrical models not requiring high mantle conductivities, conspire to make it less likely than previously supposed.

Steven Constable
Institute of Geophysics and Planetary Physics,
University of California, San Diego,
La Jolla, California 92093-0225, USA

1. Karato, S. *Nature* **347**, 272–273 (1990).
2. Oldenburg, D. W. *Geophys. J. R. astr. Soc.* **65**, 359–394 (1981).
3. Constable, S. C., Shankland, T. J. & Duba, A. *J. geophys. Res.* **97**, 3397–3404 (1992).
4. Heinson, G. & Constable, S. C. *Geophys. J. Int.* **110**, 159–179 (1992).
5. Mackwell, S. J. & Kohlstedt, D. L. *J. geophys. Res.* **95**, 5079–5088 (1990).
6. Bell, D. R. & Rossman, G. R. *Science* **255**, 1391–1397 (1992).
7. Shankland, T. J. & Duba, A. G. *Geophys. J. Int.* **103**, 25–31 (1990).
8. Bai, Q. & Kohlstedt, D. L. *Nature* **357**, 672–674 (1992).

Rewriting history

Michael D. Coe

Mesoamerican Writing Systems: Propaganda, Myth, and History in Four Ancient Civilizations. By Joyce Marcus. Princeton University Press: 1993. Pp. 495. \$49.95, £35.

In spite of its title, this book is not about Mesoamerican writing systems *per se*, but about the anthropological context in which they rose and flourished, and the role that they played in four Mesoamerican societies: Maya, Zapotec, Mixtec and Aztec. The author's main thesis is based on a distinction that she makes between history, "true events set out in a chronological framework"; myth, "a romantic and ancient falsehood"; and propaganda, a form of deliberate mind control. According to Marcus, writing came about in Mesoamerica both as a tool and as a by-product of the competition for prestige and leadership in nascent states, and remained a "propaganda tool of the state". There are two kinds of propaganda: vertical, by which rulers attempt to influence the behaviour of the ruled, and horizontal, in which elite members try to influence each other. Thus, what passes for 'history' in Maya monumental texts is really nothing more than vertical propaganda on a massive scale, so that recent efforts to write a "history in our sense" of the Classic Maya (meaning by this the barely mentioned *A Forest of Kings* by Linda Schele and David Freidel; Morrow, 1990) are doomed to failure.

There is little that is new in the first part of this thesis: it has been accepted for decades by epigraphers that the most ancient writing in Mesoamerica — Zapotec, Izapan and early Maya — appeared in a context that celebrated the doings of probably upstart rulers. As in ancient Egypt and China, but in contrast to Sumer (where the script was a vehicle for economic accounts), Mesoamerican writing became and remained an expression of royal power.

Because Marcus makes little distinction between these four scripts as systems, she tends to view the Maya case in terms of what we know about Mixtec and Aztec writing. If 'writing' is defined as the visual expression of utterances in a particular language, then these latter two scripts barely meet this criterion: in the four pre-Hispanic Mixtec codices, and in post-Conquest Aztec

documents, much of the information is pictorial, with calendrical notations given in the cyclical 52-year count. Beyond this, all we have are place names based in part on the rebus ('puzzle-writing') principle, and personal names



The mythical first human couple, Oxomoco and Cipactonal, as portrayed in the sixteenth-century Aztec Codex Borbonicus. They are seen here letting blood and casting lots. The picture is taken from the cover of *The Gods and Symbols of Ancient Mexico and the Maya: An Illustrated Dictionary of Mesoamerican Religion* by Mary Miller and Karl Taube. Thames and Hudson, £16.95.

that are largely calendrical and/or rebus. Here, the hypothesis of vertical propaganda may have some foundation, for the paucity of the linguistic information and the cyclicity of the time dimension give ample scope to manipulation of 'history'. With the Zapotec inscriptions of Oaxaca, which remain undeciphered except for the calendrical notations, the propaganda-versus-history hypothesis cannot yet be tested, despite Marcus's assertions to the contrary.

With Maya writing, Marcus is on very shaky ground. In Classic times, events were placed on a consecutive, day-to-day timescale, so that cyclicity was not a problem; moreover, the logosyllabic script was completely adequate to express any utterance in the language. Marcus's case for Classic Maya texts being unreliable propaganda rests on

two inscriptions from Palenque. One is the text on the sarcophagus in the Temple of the Inscriptions, which states that the ruler Pacal (who was buried in the tomb) died at the advanced age of 80. Physical anthropologists 'proved' that the remains were those of a man of about 40 to 50 years old. *Ergo*, the text is lying. Yet this osteological study was a travesty of scientific method: some of Pacal's bones were even lost in the process. The second is the very long text in the Temple of the Cross; Marcus makes much of the fact that it begins with the birth of ancestors who lived to Methuselan ages in an epoch thousands

of years ago. If the scribes could make such claims, then we should not rely on any 'historical' data from later times. But Floyd Lounsbury showed in 1976 that these ancient dates were contrived so that one could count back from events in the life of the current ruler (Chan Bahlum) in multiples of important astronomical cycles; this has no bearing on the historicity of Chan Bahlum or his own dates.

There are even more serious flaws in this book. While purporting to be about writing systems, it neglects two recent studies that supersede previous work on this subject: Geoffrey Sampson's *Writing Systems* (Hutchinson, 1985) and John DeFrancis's *Visible Speech* (University of Hawaii Press, 1989), both by linguists; in fact, linguistics gets the cold shoulder throughout. More disturbing is the distorted history of the Maya decipherment offered here. The contribution of the Russian Y. V. Knorosov to the decipherment of the phonetic element in the script is passed over, Marcus making it seem that the great

breakthrough was the 1962 *Catalog* published by Eric Thompson — when, in fact, it was Thompson himself who blocked the decipherment for more than three decades and who led the unjust attack on Knorosov.

One would never know from this book about the revolution in our knowledge of the Maya script that has come about through the work of such scholars as David Kelley, Linda Schele, David Stuart, Peter Mathews and Stephen Houston. The brilliant young German epigrapher Nikolai Grube is missing from the bibliography, even though 41 of the author's own works appear there (for that matter, Gordon Whittaker and Javier Urcid get the same treatment for the Zapotec script). Text citations, usually without page numbers, are stacked up in such a way that they —

along with misleading picture captions — lead one to false conclusions about who made a particular decipherment.

There is no mention of ceramics, even though the bulk of Maya texts are on pottery; this, in spite of much recent research that has cast light on the position of Maya scribes, on shamanism and beliefs about the underworld, and on the relation of scenes and texts to the mythic world of the *Popol Vuh*, the great epic of the Quiché Mayans.

The book is therefore hopelessly out of date, a conclusion strengthened by the inept drawings of Maya glyphs, which revert to a style prevalent before the publication of Catherwood's renderings in 1841. A colleague has described this as a 'Rip Van Winkle book'. Yet even Rip, waking from his 20-year slumber, finally realized that a revolution had taken place. Sadly, King George III is alive and well in these pages. □

Michael D. Coe is at the Peabody Museum, Yale University, New Haven, Connecticut 06511-8161, USA.

■ Newly published in paperback is *The Mesoamerican Ballgame* edited by V. L. Scarborough and D. R. Wilcox. Played with a solid rubber ball on masonry courts, this sport extended all over pre-Hispanic Mexico and Central America. The book claims to provide "complete coverage of the archaeological, sociopolitical, iconographic and ideological aspects of the game". University of Arizona Press, \$18.95.

Science on trial

Lee Loevinger

Regulating Toxic Substances: A Philosophy of Science and the Law. By Carl F. Cranor. Oxford University Press: 1993. Pp. 252. \$45, £35.

THOSE of us who have laboured long in the area where law and science intersect have generally agreed that an important concern is to make law more responsive to science. Now comes Carl Cranor, a philosopher who is neither a scientist nor a lawyer, to tell us that law is too responsible to science. His thesis is that "we should avoid the temptation to adopt the ideals of research science in torts and administrative law because this tends to lead to an excess of false negatives and to underregulation of carcinogens as well as to the undercompensation of plaintiffs in tort cases".

The book's title is broader than the text, as the book concerns only carcinogens, rather than toxic substances generally, and its science is restricted to epidemiology. Aetiology is mentioned only to state that the aetiology of car-

cinogenesis is not understood. All potential defects of animal bioassays and sampling errors are explored at length; and extensive statistical analysis purports to show that risk assessment is unreliable. Cranor explains that standard statistical analysis requires acceptable data to have a probability of a type I error, or false positive, no greater than 0.05, which means that researchers can have 95 per cent confidence in their results. He correctly observes that this is a common confidence level in science.

Cranor argues that the 95 per cent rule is appropriate in scientific research, but not in law because we must tolerate false positives (false diagnoses of cancer) resulting from a lower standard of proof to reduce false negatives, thus providing maximum legal protection (against cancer). The practical justification for this position is the economic assumption that regulatory false negatives are ten times as costly as regulatory false positives. The philosophical argument is that any standard of proof implies normative judgement, and it is better to overregulate than to underregulate cancer hazards.

One adventitious circumstance may make much of the discussion of tort cases irrelevant. This book apparently went to press in December 1991. But in October 1992 the US Supreme Court took jurisdiction of a case (*Daubert v. Merrell Dow* 92-102) in which the issue is what criteria courts should use in determining the admissibility of scientific evidence. The case should be decided by the end of June. While no one can predict the decision, it will probably establish definitive rules for consideration of scientific evidence in tort cases.

Whatever the decision, it will surely be based on appraisal of competing normative or policy considerations. Categories of proof in different types of cases are based on judicial judgements of relative risks and values. Although not expressed with statistical precision, these distil much practical experience which may be as valuable as philosophical rumination.

Most of this book is polemical rather than philosophical. In the last chapter, Cranor rejects utilitarianism and advocates an approach based on an artificial and unrealistic hypothesis that is unimaginable except to the mind of a philos-



ECO-DRIVE — tourists in the United States were familiar with Yosemite Valley more than a decade before the Grand Canyon and Yellowstone were even explored. This 1920 advertisement for Horseshoe Route appears in *Yosemite: The Embattled Wilderness* by Alfred Runte, which addresses the conflict between preservation and use of America's national parks. University of Nebraska Press, \$14.95, £12.95 (pbk).

opher. Neither Jeremy Bentham's utilitarianism nor Rawls's theory of justice is influential in modern law. Cranor is flogging a dead horse.

Cranor seems to lack a feel for the flexibility of Anglo-American common law and to misunderstand the significance of the categorical standards of legal proof. For example, he quotes the aphorism that in criminal law it is regarded as better that ten guilty men go free than that one innocent man be wrongly convicted. This, he says, implies that in criminal law, false negatives should be in the ratio of 10:1 to false positives. This is absurd. Every rational judge would be outraged at the suggestion that ten per cent of convictions are false positives, or convictions of innocent people. Although there are rare cases in which innocent people are convicted, the percentage is tiny.

In short, despite professions to the contrary, Cranor is anti-science, as exemplified by his statement that the effort to make risk assessment more scientific is "wrongheaded" (page 129). He correctly asserts that science is no substitute for morals. But neither is philosophy a substitute for science.

Plaintiffs' lawyers in cancer cases will find this work most helpful in discrediting scientific evidence used against their claims. Others will find it less interesting. □

Lee Loevinger is at Hogan & Hartson, 555 Thirteenth Street NW, Washington DC 20004-1109, USA.

Caveat emptor!

John Postgate

Encyclopedia of Microbiology. Edited by Joshua Lederberg. *Academic: 1992.* Four volumes (pp. 648, 646, 642 and 582). £450, \$695.

THIS is not an encyclopaedia at all. By definition — and also according to bibliographic usage — an encyclopaedia must be complete: a comprehensive repository of learning, even if today some restriction on the disciplines covered is acceptable. Complete coverage is not attempted here. Although the preface declares that the work's objective is to "survey the entire field [of microbiology] coherently" it then, contradicting itself, describes its contents as "complementing material that would be included in . . . university study".

The work comprises more than 200 review-type articles on diverse aspects of microbiology, intended to be "accessible to talented high school and college students" as well as to teachers, technicians and research specialists. The articles do not include reference citations, but end with recommendations for further reading. Useful adjuncts are a preface and guide to microbiological journals (repeated at the beginning of each volume); a glossary of relevant specialized terms (at the head of each article); and a list of contributors, a regular index, a cross-index bringing related titles together, and linkage maps of *Bacillus subtilis*, *Escherichia coli* and *Salmonella typhimurium* (all at the end of Volume 4). The two-column layout is easy on the eye; diagrams are good, but reproduction of photographs is so-so.

The variety of topics covered by the contributors is indeed eclectic; here is a selection: acetogens, biomonitors, catabolite repression, ELISA, food biotechnology, Gram-positive cocci, hepatitis, interferons, laboratory safety, "Method, philosophy of", nucleotide metabolism, origin of life, phylogeny, *recA*, sulphur cycle, T cells, ureases, new viruses, wine. Obviously it would take an unconscionable time to assess each of the 200 or so articles critically, and I have not attempted to do so, although I have browsed through most and read some 70 carefully.

Their accuracy, as far as I can judge, is generally impeccable, and the level of clarity is commendably high: I had no serious difficulty with wholly unfamiliar aspects of the subject. Indeed, a few, such as those on influenza and ribosomes, are models of presentation. The majority are useful, workman-like surveys, but several are too focused on the United States to be of more than

parochial use (such as those on careers in microbiology, guidelines for genetic manipulation, and, especially inexcusable in view of its African dimension, AIDS). One or two are on the borderlines of microbiology (plant cell culture, vertebrate tissue specificity). Some articles are more verbose and convoluted than others, and there are those that irritate the specialist because of strange omissions or emphases — but unevenness of that kind is to be expected in a multi-authored work.

I found few misprints and misspellings, but some evidence of indifferent editing. One is inured to 'data' and 'none' appearing indiscriminately as singular or plural nouns, but what editor could allow the contributor on bacterial adhesion consistently to use "flagellae" as the plural of flagellum? (Happily the definitive article on flagella gets it right.) A conflict between the equation for nitrogenase action under 'nitrogen cycle' and that in the very next article, on nodules, ought to have been resolved (the second is correct). And some contributors have been so accustomed to using reference citations in their texts that they have not adjusted to their absence, so we find, in one piece, a useless allusion to an anonymous "the same author"; in another, mention of the views of one Brian Spratt, with no further provenance.

However, the work's most serious fault, given its title, remains the lack of encyclopaedic coverage. For example, the only article specifically about antibiotics deals with microbial resistance to them, and although they do turn up in other contexts (under antifungals, ribosomes, biopharmaceuticals and so on), there are no coherent accounts of their production, properties, modes of action, biological significance and requirement by certain mutants. Again, the article on chemotaxis is first-class, but phototaxis, magnetotaxis and less-studied tactic responses are not dealt with. And yet again, the physiology of starvation is wholly ignored, in favour of admittedly fascinating recent work on adriamycin stress proteins. These are but three of many gaps I noticed.

Information transfer in science is approaching a major discontinuity, as the electronic revolution of the 1980s develops and busy researchers cease to bother with anything that is too old to have got into the computer databases. Therefore, if only as a springboard for an on-line future, a proper encyclopaedia of microbiology would have been of tremendous value. This is no such thing. Yet regarded as a grand compendium of essay-reviews dealing with aspects of modern microbiology, the book is a wide-ranging and often absorbing companion to regular textbook mat-

erial, especially for US nationals. Why, then, has it been so misleadingly named? I trust it is simply that the editorial board was muddled, as the contradiction in the preface suggests, and not a matter of commercial chicanery. □

John Postgate is emeritus professor of microbiology at the University of Sussex, Brighton BN1 9RQ, UK.

Childhood matters

Andrew Whiten

Beyond Modularity: A Developmental Perspective on Cognitive Science. By Annette Karmiloff-Smith. *MIT Press: 1992.* Pp. 234. \$34.95, £27.50.

"I HATE babies!" the cognitive scientist inside the author's head wants to shout, railing against the tribute paid to one of her lectures by a student clearly swayed by the cuteness of her subject — the human infant. Karmiloff-Smith goes on to reassure us that such irritation is not what she feels as a parent, but the welling outburst illustrates graphically a central intention of the book: "to treat cognitive development as a serious theoretical science contributing to the discussion of *how* the human mind is organised internally, and not as merely a cute empirical database about *when* external behaviour can be observed."

The study of development is thus a means to an end, rather than an end in itself. It is advocated as a tool with a unique part to play in establishing the nature of the human mind. In recent years, the computational techniques that help to define 'cognitive science' have been directed at understanding aspects of developmental psychology: here, Karmiloff-Smith aims to turn the tables and explore what developmental research can contribute to cognitive science as a whole. The extent to which she can be judged to have succeeded depends on the scope of 'human mind' that developmental study is claimed to illuminate.

If this is the adult mind, as Karmiloff-Smith implies in some parts of the text, then the significance of development must lie solely in its legacy to this mature state. The adult mind will have such-and-such an architecture *because* the developmental processes used to build it are of a certain character. This is an ambitious claim and the data marshalled here explicitly to support it seem to be relatively few so far, although clearly of great theoretical importance (E. Spelke's work on object perception

is an example).

If, on the other hand, 'human mind' is not restricted to adulthood, then it might be thought that claiming an important role for developmental study is little more than a truism. But this neglects the fact that influential theories in contemporary cognitive science have been elaborated as if development did not exist. At its most simplistic, this has involved the writing-off of ontogenetic processes by labelling certain cognitive systems 'innate', a term abandoned 20 years ago by many workers in the subject that had earlier spawned it, development ethology.

But the mind does develop: and to me the success of this book lies not so much in the author's ambition of demonstrating the relevance of development for understanding any mental end-state, but rather in her impressive developmental analysis *per se*, which deserves wide readership by both developmentalists and nondevelopmentalists who need an overview of the state of the art. Clearly and comprehensively, Karmiloff-Smith shows the highly structured ways in which different representational processes emerge from infancy onwards. It is the totality of these structures that should surely constitute the subject matter of cognitive science.

Successive chapters deal with several principal areas of knowledge, examining the child in turn as linguist, physicist, mathematician, psychologist and notator. In themselves these offer excellent, up-to-date, but fairly concise reviews of current knowledge in each field. Repeatedly one finds explications of difficult concepts, such as the propositional attitudes underlying theory of mind, that are among the clearest available.

But the author goes much further than this, offering a sophisticated model of cognitive change that distinguishes four principal representational formats (in place of the conventional implicit-explicit dichotomy), proceeding from behavioural mastery to metalinguistic awareness. These are not stages in the Piagetian sense, but rather sequential phases of representational redescription that can recur at different ages in different domains. Karmiloff-Smith takes us indeed far beyond (and nicely between) two positions: the Piagetian notion of development as one of several generalized intellectual transitions, and the notion of major cognitive systems, such as theory of mind, as innate modules. Increasingly it looks as if beyond and between these is the place we need to be. □

Andrew Whiten is in the Department of Psychology, University of St Andrews, St Andrews, Fife KY16 9JU, UK.

Pain relief

Gerald Zernig

Forschen auf Deutsch: der Machiavelli für Forscher — und solche die es noch werden wollen*. By Siegfried Bär, with cartoons by Irena Volpi. Harri Deutsch, Frankfurt: 1992. Pp. 125. DM20 (pbk).

YOU crawl home after one of those days in the lab. Your graduate student has wasted the whole digitonin batch. The manuscript that you have slaved over for the past year has been rejected — but then again, so has your grant application. You are dangerously close to the end of your contract, a fact that has just today been brought to your attention (again!) by your supervisor. To round things up, you had to spend the past hour convincing the usual salesman that you really didn't need a third centrifuge, much less could afford one. You are in terrible need of something, someone to soothe your soul.

Here it is! Written by a postdoc, the same downtrodden academic creature as yourself, this guide in German offers a wildly biased, cynical view of academic life. You feel immediately drawn to the slashing description of all that oppresses you, your neck aches from nodding in bitter appreciation, sometimes you find yourself rolling on the floor with laughter. You have found a voice, someone exposing all the injustices that have befallen you. You cheer, you agree wholeheartedly, and you quickly skip those pages where the author offers his own ambitious advice for solving your problems (after all, you don't want to spoil that sweet feeling of complete harmony with the author).

Then, after a while, with composure regained, you are able to take a look at yourself, your own failings and shortcomings, your own efforts that, this very evening, look so hopeless to you. So many refreshing passages convince you that you are not alone. For example:

Usually, the researcher reads only articles of his very own field in their entirety. Nevertheless he spends a considerable part of his time copying papers which, mostly unread, are stacked or filed. For our time-pressed researcher, copying represents spiritual ownership, an act taking on the importance of a ritual that replaces the tedious process of actually reading the paper. Copying soothes the researcher and blesses him with inner peace and the feeling not to have missed an important bit of information.

There are lots of cartoons, ideally suited for the department coffee room, depicting, for example, the Measure-

ments Flunky (his head "contains raisins in a cake mix of little material expectations, ambition and *naïveté*") or a chart on Academic Metabolism (with the inevitable "USA shunt"). There are interesting bits of information snatched from *Nature* and *Science* on science politics: 20 per cent of all biochemical or molecular biology articles — or 40 per cent of all medical papers, for that matter — are never cited. And how about this: in 1982, the average German research paper (which cost the taxpayer \$188,570) was cited only 0.85 times, whereas the average British one could boast an equally meagre 0.96 citations (while, admittedly, costing the UK citizen only \$81,260). Sobering facts for all, whether graduates, postdocs, research group leaders, chairmen, deans or government science administrators.

This is a book that you will want to give your child to prevent it from foolishly entering academic science as you yourself did. And it is just the book you need to hang on to until you are rewarded, again, with that exhilarating feeling that, after all, science is the "most exciting thing you can do with your pants on", as you will hear — and wholeheartedly agree with — during the late hours of the banquet of your next congress, when all will have ended well. □

Gerald Zernig is in the Department of Pharmacology, University of Michigan Medical School, 6322 Medical Science Building I, Ann Arbor, Michigan 48109-0626, USA.

New Journals Issue

This year, *Nature's* annual new journals review supplement will appear in the issue of 7 October. Publishers and learned societies are invited to submit journals for review, taking note of the following criteria:

■ Journals that first appeared during or after June 1991 and issued at least four separate numbers by the end of April 1993 will be considered.

■ Journals covering any aspect of science are eligible, although those dealing with clinical medicine, engineering and pure mathematics are excluded, as are publications of abstracts.

■ Frequency of publication must be at least three times a year. The main language used must be English. Translation journals in English are, of course, eligible.

■ Deadline for submission is the end of May.

When submitting journals for review, please send at least four different issues (the first, the most recent and any two others) of each title, together with full details of subscription rates (personal and institutional) and frequency of publication, to: Peter Tallack, *Nature*, 4 Little Essex Street, London WC2R 3LF, UK. For further information please telephone Peter Tallack on 071-836-6633 (011-44-71-836-6633 from the United States), extension 2414.

*Researching in German: The Machiavelli for scientists — and those who still want to become one.

Instability and decay of the primary structure of DNA

Tomas Lindahl

Although DNA is the carrier of genetic information, it has limited chemical stability. Hydrolysis, oxidation and nonenzymatic methylation of DNA occur at significant rates *in vivo*, and are counteracted by specific DNA repair processes. The spontaneous decay of DNA is likely to be a major factor in mutagenesis, carcinogenesis and ageing, and also sets limits for the recovery of DNA fragments from fossils.

ALL biological macromolecules spontaneously decompose. Proteins that renature completely after heat-induced unfolding in neutral solution are susceptible to slow and irreversible inactivation at elevated temperatures which is independent of concentration. The intact covalent structure of the unfolded form of a typical enzyme such as egg-white lysozyme has a half-life of only a few minutes at pH 7.4 and 100 °C, largely due to deamidation of asparagine residues¹. Such simple chemical processes occur also at lower temperatures, with a typical roughly threefold decrease in reaction rate for each 10 °C interval, and deamidated asparagine residues not protected by tertiary structure have been observed in trypsin crystals after prolonged growth periods². Nucleic acids also undergo spontaneous decomposition in solution, RNA being particularly vulnerable. Because of the presence of the 2'-hydroxyl group of ribose, the phosphodiester bonds of RNA molecules are very susceptible to hydrolysis, particularly in the presence of divalent cations such as Mg²⁺ and Ca²⁺ (ref. 3). Transition metals, for example Zn²⁺, have an even greater deleterious effect. This property poses a major problem for predictions of the early existence of an aqueous 'RNA world' in connection with the origin of life⁴. Although RNA can serve as an adequate but relatively short-lived and mutation-prone carrier for a limited amount of genetic information, as seen in RNA viruses, the development of autonomously replicating cells probably depended on reduction of the ribose moiety in nucleotides to the unusual sugar, deoxyribose. This event provided genomes of greatly improved chemical stability.

DNA hydrolysis

The chemical price paid for the greatly increased resistance of the nucleic acid phosphodiester bond (gained by removal of the sugar 2'-OH group) is a labile *N*-glycosyl bond. Base-sugar bonds of ribonucleosides are much less susceptible to hydrolysis than those of deoxyribonucleosides. The instability of DNA glycosyl bonds under physiological solvent conditions was investigated 20 years ago; it was measured by using DNA with ¹⁴C-labelled purine or pyrimidine residues and determining the rates of release of free bases as a function of temperature, pH, ionic strength and nucleic acid secondary structure⁵⁻⁷. A more sensitive method involves registering the introduction of apurinic sites in covalently closed circular DNA molecules as they become sensitive to cleavage by alkali or specific DNA repair enzymes, AP endonucleases⁸. Guanine and adenine are liberated from DNA at similar rates, with guanine being released slightly more rapidly, whereas cytosine and thymine are lost at 5% of the rate of the purines. The difference in depurination velocity between single-stranded and double-stranded DNA is only four-fold, so the double helical structure does not provide much protection. As is the case for deoxyribonucleosides⁹, the reaction is mainly acid-catalysed under physiological conditions⁵. Effective charge-shielding of DNA with Mg²⁺ does not measurably influence the rates of hydrolytic release of free bases. At a site of base loss, the DNA chain is weakened and undergoes

cleavage by a β -elimination process within a few days⁸⁻¹⁰. This reflects the fact that the base-free sugar residue exists in an equilibrium between the major cyclic form and about 1% of the acyclic reactive aldehyde form.

In metabolically active cells, DNA occurs in the fully hydrated B form and would be expected to undergo depurination (with accompanying loss of genetic information) at a rate similar to that of DNA in solution. Fortunately, ubiquitous AP endonucleases rapidly initiate a DNA repair process (Fig. 1) by introducing DNA strand breaks on the 5' side of base-free sites. The sugar-phosphate residue is subsequently removed by a separate phosphodiesterase, followed by filling-in of the one-nucleotide gap by DNA polymerase and DNA ligase^{10,11}. This excision-repair pathway can handle much larger loads of DNA depurination than that produced by spontaneous hydrolysis. Thus, an *Escherichia coli* cell growing at 37 °C in the presence of a high, but nonlethal, concentration of the alkylating agent methyl methanesulphonate would suffer (and successfully repair) several thousand apurinic sites per generation, caused by release of the alkylation products 3-methyladenine and 7-methylguanine. In contrast, only a single apurinic site would emerge every second generation by spontaneous depurination of the *E. coli* chromosome. This reaction must, nevertheless, be of significance for mammalian cells in view of their very large genome size and slow replication, and it can be estimated that in each human cell 2,000-10,000 DNA purine bases turn over every day owing to hydrolytic depurination and subsequent repair⁵.

In addition to the problem of the intrinsic lability of glycosyl bonds, DNA base residues are susceptible to hydrolytic deamination. As expected from data on nucleosides⁹, cytosine and its homologue 5-methylcytosine are the main targets for this reaction. The deamination rates of these bases were defined experimentally by following their conversion to uracil or thymine residues in suitably radioactively labelled DNA¹²⁻¹⁴ as a function of temperature, pH and DNA secondary structure. Because the reaction is slow at 37 °C, the rate constants at various higher temperatures have been measured. This procedure, which yielded the activation energy of the reaction, allowed the cytosine deamination rate in DNA at 37 °C to be deduced. Recently, a very sensitive genetic reversion assay measuring the rate of deamination at a single DNA cytosine residue in the *E. coli lacZ* gene was used to reinvestigate this problem¹⁵. The cytosine deamination rate was determined directly for both single-stranded and double-stranded DNA at 37 °C and 30 °C. This genetical approach validated the extrapolation from elevated temperatures in the biochemical experiments in that the estimated cytosine deamination rates at 37 °C were the same within experimental error. For single-stranded DNA in solution, the half-life of an individual cytosine residue is about 200 years at 37 °C and pH 7.4. In such DNA, loss of purine residue occurs at a similar or slightly slower rate. In contrast to depurination, however, the double helix structure affords very good protection

against hydrolytic cytosine deamination, and this reaction occurs at only 0.5–0.7% of the rate of single-stranded DNA, that is, with a half-life of about 30,000 years for each cytosine residue. Heat-induced deamination of the 5-hydroxymethylcytosine residues found in bacteriophage T4 DNA to 5-hydroxythymine takes place at a somewhat higher rate¹⁶.

The mechanism of hydrolytic conversion of cytosine to uracil in DNA at neutral pH involves direct deamination by alkali-catalysed hydrolysis and also attack by water on the protonated base in a general acid-catalysed reaction⁹. Consequently, there is no strong pH dependence in the vicinity of pH 7.4 (ref. 12). An alternative mechanism involving an addition–elimination reaction⁹, analogous to the deamination of cytosine residues by bisulphite treatment, seems less likely because it would yield relatively long-lived pyrimidine hydrates as reaction intermediates; these are recognized by a different DNA repair enzyme than that acting at uracil residues. Furthermore, if the addition–elimination reaction occurred, for both steric and electrostatic reasons, 5-methylcytosine should be more refractory to deamination than cytosine residues. Instead, 5-methylcytosine moieties are deaminated three to four times more rapidly than cytosines^{12,14}. In a mammalian cell with 3% of its DNA cytosine residues in methylated form, about 10% of hydrolytic deamination would occur at 5-methylcytosine residues and 90% at cytosine residues. This relatively small difference in deamination rates is by itself unremarkable, and would hardly make 5-methylcytosine residues greatly preferred targets for spontaneous mutagenesis. But the effect is amplified by different rates of DNA repair. The deaminated form of cytosine is very rapidly excised by the abundant and ubiquitous uracil-DNA glycosylase to generate a base-free site, which is efficiently corrected (Fig. 1). This DNA repair enzyme cannot act on deaminated 5-methylcytosine, that is, thymine. The G·T base-pair in DNA is, instead, a substrate for mismatch correction systems, but these processes are slower than repair of G·U base pairs. A correspondingly greater risk for mutation fixation, therefore, will occur. In wild-type *E. coli*, 5-methylcytosine residues are mutational hot spots, whereas all cytosines are hot spots in a mutator *ung*[−] strain deficient in uracil-DNA glycosylase¹⁷. In agreement with these data, *Streptococcus pneumoniae* and *Saccharomyces cerevisiae* strains deficient in uracil-DNA glycosylase also have a 10–20-fold increased spontaneous mutation rate caused by deamination of DNA cytosine residues^{18,19}. The hydrolytic reaction occurs without any apparent sequence specificity. Interestingly, the rate of spontaneous deamination of DNA cytosine residues is estimated to be 40-fold higher in *S. cerevisiae* than in *E. coli*, possibly because the slower rate of eukaryotic transcription might keep DNA locally in a single-stranded form for a more extended period¹⁹. The higher rate of deamination of 5-methylcytosine, combined with its relatively slow repair in mammalian cells, contribute to making methylated CpG sequences preferential (about 40-fold) targets for spontaneous point mutations; such G·C→A·T transitions account for one-third of single-site mutations observed in inherited human disease²⁰. Direct genomic sequencing of the p53 tumour suppressor gene showed that cytosine residues known to have undergone mutation were methylated in normal and tumour tissues, indicating that “5-methylcytosine functions as an endogenous mutagen and carcinogen in humans”²¹. This pattern of reduced correction of lesions arising from deamination of 5-methylcytosine may also account for the underrepresentation of CpG dinucleotides in mammalian genomes.

A separate factor contributing to the deamination of CpG sequences in mammalian cells might be ascribed to the catalytic mechanism in which cytosine is converted to 5-methylcytosine by the DNA methyltransferase²². The enzyme forms a covalent bond with the C6 position of the cytosine in CpG, thereby generating a transient dihydropyrimidine reaction intermediate that is highly susceptible to spontaneous deamination. So far, an enzyme-induced deamination reaction of this type has only

been demonstrated in cells with an *ung*[−] genotype at levels of the S-adenosylmethionine cofactor far below the physiological concentration (~40 μM), but such active deamination might also occur occasionally in normal cells²³. This could explain the finding that apparently unmethylated CpG doublets in α-globin gene sequences exhibit an increased mutation frequency²⁴. Furthermore, enzyme-promoted deamination provides an attractive explanation for the very rapid alteration of extensively methylated, recently duplicated DNA sequences in *Neurospora crassa*, designated repeat induced point mutations (RIP, ref. 22).

By comparison with the hydrolytic conversion of cytosine to uracil, the deamination of DNA purines is a minor reaction. Adenine is converted to hypoxanthine (the base residue in deoxyinosine) in DNA at 2–3% of the rate of cytosine deamination²⁵. Because this moiety forms a more stable base pair with C than with T, the product of hydrolytic deamination of adenine is a mutagenic lesion. It appears to be repaired in an analogous fashion to uracil in DNA but by a separate DNA glycosylase²⁶. The repair reaction is, however, less efficient because of the low cellular level of hypoxanthine-DNA glycosylase²⁷. The rate of deamination of guanine to xanthine in DNA has not been precisely determined but is similar to, or slower than, adenine deamination. No specific repair enzyme for xanthine residues in DNA has been identified. Both hypoxanthine and xanthine base-pair preferentially with cytosine, albeit with reduced coding specificity, so rare deamination of adenine residues may be much more deleterious than that of guanine from the point of view of generating premutagenic lesions. In addition, the xanthine-deoxyribose bond is particularly susceptible to spontaneous hydrolysis, indicating that a dXMP residue generated by a rare guanine deamination event in a mammalian cell would be converted to an apurinic site within days or weeks at 37 °C in a nonenzymatic reaction. In view of the relatively poor cellular response mounted to deaminated purines compared to uracil residues, nitrosative deamination of DNA is unlikely to have been of general importance during evolution, because this process acts similarly on guanine, adenine and cytosine. A recent proposal suggesting that ‘nitrosative deamination of DNA by nitric oxide may represent an important mechanism of genomic alteration’²⁸ therefore seems unlikely. Possibly, endogenously produced nitric oxide could be of occasional mutagenic relevance in organs exposed to large amounts of this short-lived signal compound.

Other types of hydrolytic damage may conceivably result in lesions in DNA, but there is no biochemical evidence for such reactions. A mutagenic guanine derivative resulting in G→T transversions was observed in phage T4 DNA after prolonged incubation, and it was speculated that this might have resulted from heat-induced glycosyl bond migration¹⁶. Alternatively, this event might reflect oxidative formation of 8-hydroxyguanine residues.

DNA oxidation

Aerobically growing cells are exposed to active oxygen during normal metabolism. This represents an important source of endogenous damage to the genome. Bacteria possess elaborate inducible responses to oxidative stress, including the overproduction of DNA repair enzymes that remove blocked 3' termini at strand breaks²⁹. The major mutagenic base lesion generated by hydroxyl radicals is 8-hydroxyguanine, which base-pairs preferentially with adenine rather than cytosine and thus generates transversion mutations after replication^{30,31}. This lesion is excised by a specific DNA glycosylase, present both in *E. coli* and in mammalian cells, which also removes other altered purine residues with ring-opened rather than oxidized imidazole moieties, that is formamidopyrimidines³². The 8-hydroxyguanine and formamidopyrimidine base derivatives are generated in similar amounts by hydroxyl and superoxide radicals³³, but 8-hydroxyguanine is of greater interest because it has a directly mutagenic effect. Inactivation of the *E. coli fpg* (*mutM*)

gene encoding the relevant DNA glycosylase leads to a 10-fold increase in spontaneous mutation frequency³⁴. The same relative increase is seen in *E. coli ung* mutants that are unable to excise uracil from DNA. It seems likely that hydrolytic deamination of cytosine to uracil in DNA, and oxidation of guanine residues to 8-hydroxyguanine, are the two major types of spontaneous, directly premutagenic events in living cells. Furthermore, they occur at similar frequencies. (By comparison, apurinic sites are mainly cytotoxic lesions that block DNA polymerases.) The data on the rate of cytosine deamination in double-stranded DNA suggest that each type of event takes place 100–500 times per day in a human cell. Oxidation of the guanine moiety of dGTP poses another serious mutagenic threat to cells, because

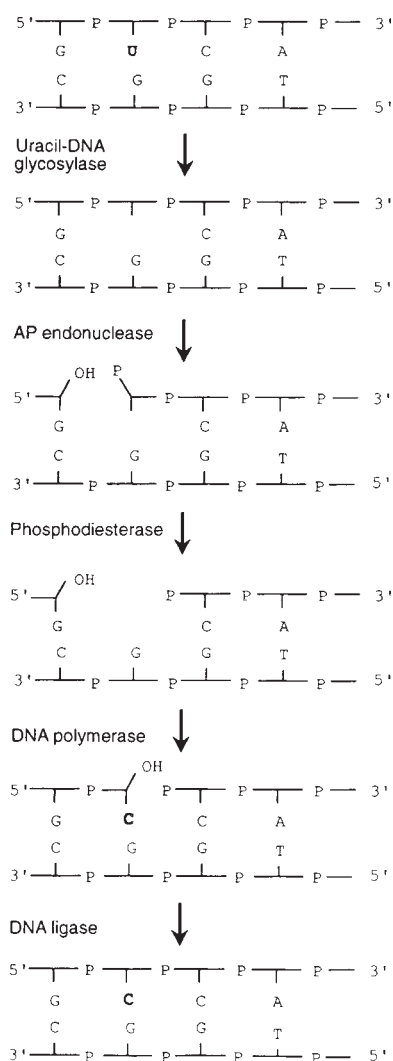


FIG. 1 The base excision-repair pathway for removal of endogenous damage from cellular DNA. The scheme is a minor modification of a previous model⁸². Apurinic/aprimidinic (AP) sites in DNA can be introduced by nonenzymatic hydrolysis of base-sugar bonds or by the action of several DNA glycosylases, specific for various types of DNA damage. Shown here is the excision of a deaminated cytosine residue (uracil) and the base-free deoxyribose-phosphate moiety, followed by replacement and correction of the missing nucleotide residue by DNA repair synthesis. Other forms of base damage generated by hydrolysis or by nonenzymatic methylation are recognized by separate DNA glycosylases, but the subsequent steps are identical. Two DNA glycosylases that counteract oxidative damage have an associated AP lyase chain-breaking activity and thus may be able to sidestep the requirement for an AP endonuclease. Attempts to construct an *E. coli* strain unable to incise DNA at AP sites have been unsuccessful, strongly indicating that such a cell would not be viable⁸³ and illustrating the importance of this DNA repair pathway.

8-hydroxy-dGMP residues can be misincorporated opposite adenine residues during DNA replication. This replication problem is dealt with by rapid degradation of 8-hydroxy-dGTP to a monophosphate by the small MutT protein, thus preventing the use of the triphosphate as a precursor for DNA synthesis³⁵. Moreover, any adenine residue that is misincorporated opposite an 8-hydroxyguanine in template DNA is excised by a specific DNA glycosylase, the product of the *E. coli mutY* gene³⁴. The cellular dUTP pool is also kept very small by a specific dUTPase. But the occasional use of dUTP instead of TTP as a precursor for DNA synthesis is harmless, because the resulting A·U base pair is equivalent to an A·T base pair. The uracil residue would in any event be rapidly removed by uracil-DNA glycosylase.

Published measurements of rates for generation of oxidative DNA damage are indirect and imprecise. Thus the estimate for 8-hydroxyguanine formation given above is 1–3 orders of magnitude lower than that in a recent proposal³⁶. The data in the latter study were of two kinds, neither of which was wholly compelling. Thus, it was reported that 8-hydroxyguanine could be detected in DNA from several rat organs. Because the lesion, being so mutagenic, is effectively repaired, it is unlikely that it would accumulate in nuclear DNA to a measurable degree, so this result may be attributable to minor oxidation of the DNA during isolation with solvents such as phenol and chloroform, and air-saturated buffers. Secondly, 8-hydroxydeoxyguanosine and smaller amounts of 8-hydroxyguanine were detected and measured in rat urine. It is a common misconception that such occurrence of the altered nucleoside reflects DNA repair. In fact, the nucleoside is not a reaction product in the base excision-repair pathway initiated by DNA glycosylase. Instead, it is presumably derived by degradation of DNA from dead cells by unspecific nucleases and phosphatases, and oxidation could have occurred during passage through the kidney.

Another type of extensively studied DNA base lesion generated by oxidation damage is a ring-saturated derivative of a pyrimidine. This lesion occurs in several forms, in particular as thymine and cytosine glycols and pyrimidine hydrates. These derivatives have lost the 5, 6 double bond and, in consequence, their planar ring structures; they are noncoding bases^{37,38}. The lesions may decompose further into fragments of bases. *In vitro*, ring-saturated pyrimidines are formed in DNA at rates marginally higher than that of 8-hydroxyguanine. They are excised by a specific DNA glycosylase with an associated AP lyase activity that promotes cleavage of the sugar-phosphate chain by β -elimination. This enzyme, which in *E. coli* has received the erroneous name 'endonuclease III', is an iron-sulphur protein; it is the first of the ubiquitous DNA repair enzymes for which the three-dimensional structure has been determined³⁹. Attempts to estimate the rates of conversion of DNA pyrimidines to ring-saturated forms *in vivo*⁴⁰ meet with the same difficulties as encountered with regard to 8-hydroxyguanine.

A particularly interesting form of oxygen radical-induced DNA damage which has received relatively little attention is that causing major helical distortion. Such lesions include 8, 5' cyclopurine deoxyribonucleosides generated by an additional covalent bond, formed between the base and the sugar-phosphate backbone⁴¹, and covalent intrastrand purine dimers⁴². These are unlikely to be substrates for DNA glycosylases and, therefore, would not be expected to be removed by the base excision-repair pathway. Recently, a major oxygen radical-induced DNA lesion of this type was found to remain unrepaired in biochemical experiments using cell extracts of xeroderma pigmentosum origin, whereas control extracts were repair-competent⁸⁴. Xeroderma pigmentosum cells are specifically defective in nucleotide excision-repair and normal with regard to base excision-repair.

Stimulated polymorphonuclear leukocytes and monocytes, and human tumour cell lines, produce substantial amounts of H₂O₂ (ref. 43). This might result in oxidative DNA damage and resulting death in adjacent cells, and chromosome instability in

tumour cells, in spite of cellular catalase activity. Some of the peroxide-generated DNA damage may proceed through a non-radical pathway to generate distinct DNA lesions, for example adenine-N1-oxide⁴⁴. The relative significance of such events compared to more abundant hydroxyl radical-induced DNA damage remains to be investigated.

The cell nucleus is a very poorly oxygenated intracellular compartment which lacks detectable O₂ metabolism⁴⁵. The delegation of the metabolism of reactive oxygen to mitochondria must protect nuclear DNA against oxidative damage; this may be a major reason for the evolution of distinct nuclei in eukaryotic cells. A corollary to this is that mitochondrial DNA should be uniquely sensitive to such damage. The situation is further aggravated by the fact that mitochondria do not contain histones which quench the generation of oxygen radical-inflicted DNA lesions⁴⁶. Thus, DNA from mitochondria contains higher levels of 8-hydroxyguanine than DNA isolated from cell nuclei. Interestingly, the amounts of this lesion are even higher in DNA from isolated mitochondria exposed to oxidative stress⁴⁷. Oxygen radical-induced DNA damage in mitochondria could be largely responsible for the high mutation rate, rapid evolution and decline with age of this organelle, as well as contributing to several maternally inherited human degenerative diseases, such as Leber's hereditary optic neuropathy^{48,49}.

Nonenzymatic DNA methylation

In addition to oxygen, living cells contain several other small reactive molecules that might cause DNA damage and act as endogenous genotoxic agents⁵⁰. The best characterized and probably most important of these is *S*-adenosylmethionine (SAM). This efficient methyl group donor is used as cofactor in most cellular transmethylation reactions. As for other simple methylsulphonium compounds, nonenzymatic transmethylation from SAM to various nucleophiles occurs at a slow rate. Thus, SAM is a weak alkylating agent and its nonenzymatic methylation of proteins⁵¹ and DNA^{52,53} can readily be demonstrated by using ³H-methyl-labelled SAM. The reaction is analogous to that of simple alkylating agents (such as methyl methanesulphonate or dimethyl sulphate) with DNA, the main targets being ring nitrogens of purine residues. 7-Methylguanine and 3-methyladenine are the major DNA lesions (Fig. 2). The former compound seems relatively harmless, because this guanine modification does not alter the coding specificity of the base. By contrast, 3-methyladenine is a cytotoxic DNA lesion that blocks replication, so it poses a major threat to the cell. It may be estimated that about 600 3-methyladenine residues per day are generated in DNA of a human cell in this reaction⁵², an alkylation level that would be achieved by continuous cellular exposure to 20 nM methyl methanesulphonate. There is no obvious protection against this DNA modification by cellular compartmentalization, because SAM is present in the cell nucleus and serves as cofactor for enzymatic DNA methylation. But both eukaryotic cells and bacteria contain a highly efficient 3-methyladenine-DNA glycosylase that rapidly excises the altered base, generating an apurinic site. 7-Methylguanine, by contrast, is poorly repaired and might be expected to accumulate to a measurable degree in the DNA of mammalian cells, although the chemical lability of the 7-methylguanine-deoxyribose bond ensures that a steady state of base modification and loss would be achieved within a few days. In agreement with this notion, direct measurements of 7-methylguanine adducts in rat liver DNA by a sensitive electrochemical detection procedure demonstrated the presence of small but significant amounts of the alkylation product⁵⁴.

Enzymatic methylation of DNA is widespread and is an important mechanism for modulating gene expression in mammalian cells. Such methylation occurs at the C5 position of cytosine, or in bacteria at the amino group of adenine. These are extremely poor acceptor sites for alkylating agents, so there is virtually no overlap between targets of enzymatic versus

nonenzymatic DNA methylation. This arrangement ensures that an enzymatically modified base residue serving as a biological signal is not accidentally removed by a specific DNA repair enzyme that counteracts alkylation damage. Similarly, the selection of thymine rather than uracil as a DNA base aids the repair of deaminated cytosine residues.

The enzyme cofactor for TMP synthesis, *N*⁵-methyltetrahydrofolic acid, has a poor transfer potential and is present at low concentrations in cells, so it is presumably much less important than SAM as a potential DNA methylating agent. Neither is likely to generate the highly mutagenic DNA lesion *O*⁶-methylguanine except in trace amounts. The production of this miscoding base accounts for the very strong mutagenic effect of alkylating agents such as *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG) and methylnitrosourea. Interestingly, *E. coli* cells specifically defective in repair of *O*⁶-methylguanine exhibit a threefold increased frequency of spontaneous mutation in stationary phase but not in logarithmic growth phase, compared to wild-type cells⁵⁵. These data imply that nongrowing *E. coli* cells may produce small amounts of an endogenous highly mutagenic alkylating compound. The identity of this agent remains unknown, but a survey of possible endogenous or environmental sources has revealed tentative candidates⁵⁶. For example, the simple metabolites carbamyl phosphate and methylamine may condense to form methylurea, which can be nitrosated to generate methylnitrosourea⁵⁷.

Many other cellular compounds, including metabolites of common amino acids, may form covalent adducts with DNA to a limited extent. This topic requires further investigation. The nonenzymatic reaction of reducing sugar compounds such as glucose-6-phosphate with amino groups of DNA bases is of particular interest⁵⁸. Under *in vivo* conditions, such adducts would only be generated at 0.1–1% of the rate of the other spontaneous DNA alterations discussed above, and the lesions would probably be repaired by the nucleotide excision-repair pathway. Nevertheless, a detectable mutagenic effect on plasmid DNA was demonstrated in *E. coli* cells with elevated glucose-6-phosphate levels⁵⁹.

Spores and thermophiles

There are biological strategies that overcome the intrinsic instability of DNA. Many microorganisms respond to hostile environmental conditions by sporulation, and DNA in some bacterial spores exhibits impressively increased resistance to heat- and radiation-induced damage. A recent review⁶⁰ summarizes the strategies used to achieve this goal. Prolonged heating of suspensions of *Bacillus subtilis* spores induces a very high mutation frequency among survivors, indicating that the introduction of lethal mutations is a major cause of loss of viability. In order to reduce hydrolytic depurination, the double-stranded DNA in spores exists in a partly dehydrated state and exhibits an A-like rather than B conformation. Moreover, a special group of small acid-soluble DNA-binding proteins is synthesized on sporulation; these proteins bind tightly and specifically to the A-form of DNA and reduce by at least 20-fold its rate of depurination. As an additional precaution, extensive DNA repair occurs during the first few minutes of spore germination. Oxygen-induced DNA damage in spores may be minimized by the DNA-binding proteins as well as by the absence of active metabolism; the lack of nucleoside triphosphates (and the low water content) in spores prevents DNA repair before germination.

In contrast to the special mechanisms devised to protect DNA in bacterial spores, there is no obvious need for unusual or exaggerated DNA repair processes in thermophilic bacteria growing at high temperatures. In fact, the efficient DNA repair mechanisms present in *E. coli* for correcting depurination and cytosine deamination would be just sufficient to withstand the roughly 3,000-fold increase in DNA decay and allow growth at 100 °C instead of 37 °C if the bacterial proteins could tolerate

the elevated temperature. But above 100–110 °C, serious difficulties would arise, both because of the chemical lability of the DNA structure to hydrolysis and the problem of retaining appropriate hydrogen-bonding between the two DNA strands. In the latter regard, high pressure affords some protection, in that the melting temperature (T_m) of the double helix is about 10 °C higher at 5,000 atmospheres. A report⁶¹ that thermophilic methanogenic bacteria are proliferating in deep-sea sulphide chimneys at temperatures above 250 °C seems amazing and is very difficult to reconcile with the known chemistry of DNA: the molecule would decompose so rapidly that no known DNA repair systems could conceivably counteract all the damage, and the preservation of the double helix would require stabilizing factors with a capacity far beyond any known compounds. Subsequent studies on this phenomenon indicate that the apparent bacterial growth at 250 °C may have been due to a chemical artefact produced at high temperatures rather than to cellular proliferation⁶².

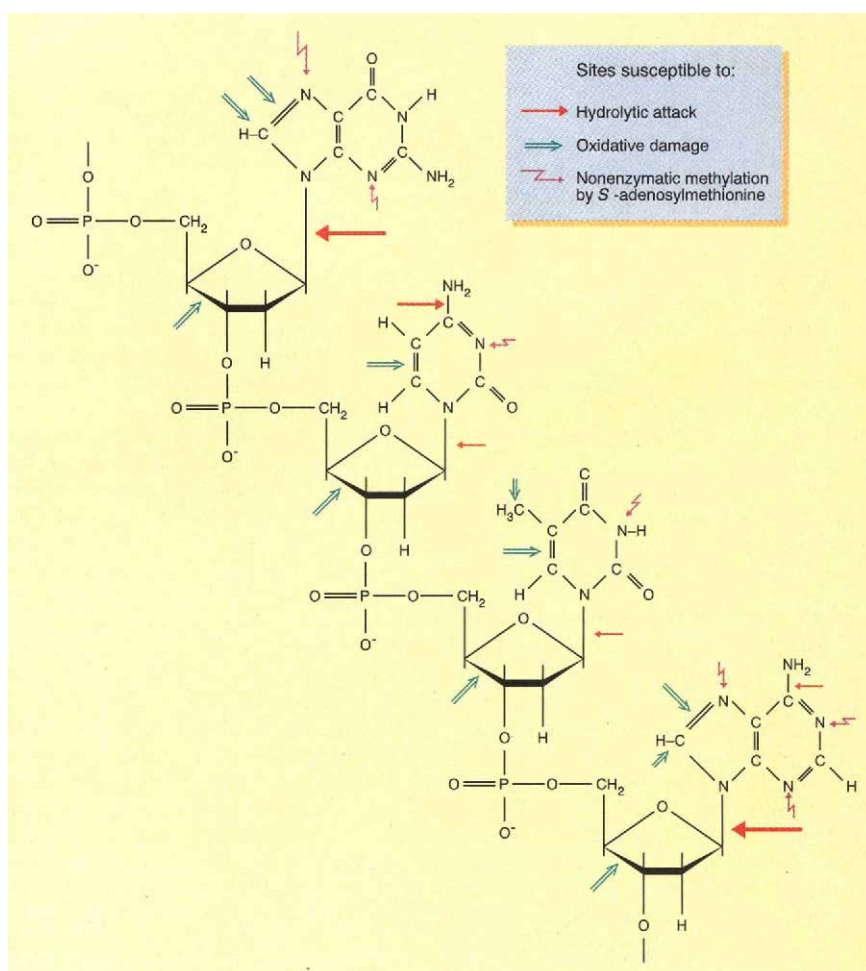
Ancient DNA

Recovery of DNA fragments from extinct animals or plants in museum collections or from archaeological excavations can permit direct comparison with related contemporary material by DNA sequencing. This approach offers a valuable complement to taxonomic studies. Short mitochondrial DNA sequences from the extinct quagga have been cloned and compared with homologous sequences⁶³ from zebra and other related animals. Similarly, 2,400-year-old DNA fragments from an Egyptian mummy have been cloned and characterized⁶⁴. The advent of the polymerase chain reaction (PCR) has greatly aided such investigations because minute amounts of short DNA fragments

can be selectively amplified⁶⁵. Thus, DNA samples from archaeological bones up to 5,000 years old have been recovered and analysed⁶⁶. A problem with the hypersensitive PCR technique, as encountered by most newcomers to the method, is that trace amounts of contaminating DNA accidentally derived from laboratory glassware, or even the experimenter, readily produce false-positive results. This is a particularly difficult problem in work with ancient DNA, which is often highly degraded or possibly non-existent in available specimens and the occasional dramatic 'success' in this area should be viewed with skepticism.

From the data on hydrolytic and oxidative decomposition reviewed above, it can be predicted that deprived of the repair mechanisms provided in living cells, fully hydrated DNA is spontaneously degraded to short fragments over a time period of several thousand years at moderate temperatures. The most important route of decay for hydrated DNA is depurination. A 5–10-fold reduction in the rate of this process can be achieved at very high ionic strength⁵. Furthermore, adsorption of DNA to hydroxyapatite results in a twofold decrease in the rate of depurination (P. Robins and T.L., unpublished data); this could slightly improve the chances of recovery of useful DNA from old bones. Thus, in connection with favourable preservation conditions, it seems feasible that useful DNA sequences tens of thousands of years old could be recovered, particularly if the fossil has been retained at low temperature. Preliminary findings have been made on the identification of short mitochondrial DNA sequences from 40,000-year-old mammoth tissue (R. Higuchi, unpublished data). The data on bacterial spores would suggest that further increased stabilization of the DNA in fossils would be achieved by partial dehydration, and by the exclusion of oxygen. Note in this regard that DNA in air-dried tissues

FIG. 2 Target sites for intracellular DNA decay. A short segment of one strand of the DNA double-helix is shown with the four common bases (from top: guanine, cytosine, thymine, adenine). Sites susceptible to hydrolytic attack are indicated by solid red arrows, oxidative damage by open green arrows, and nonenzymatic methylation by *S*-adenosylmethionine as zig-zagged purple arrows. Major sites of damage are indicated by the large arrows. Hydrolytic and oxidative damage, but not methylated residues would accumulate in fossil DNA.



remains partly hydrated and is still susceptible to decay. Because some of the water molecules in the grooves of the DNA double helix are structurally essential, 'dry DNA' obtained by storing DNA fibres over the efficient drying agent phosphorous pentoxide does not retain a double-helical conformation, making the bases more vulnerable to damage; such DNA is extremely hygroscopic and is rapidly rehydrated on exposure to air.

Recently, much excitement has been generated by claims of PCR amplification of DNA tens of millions of years old. The most intriguing of these reports concerns DNA from a 25–30-million-years-old termite preserved in amber⁶⁷. The authors tried with commendable care to avoid the contamination artefacts that have marred much of the work in this novel area of research. Minute amounts of highly degraded DNA were extracted from the preserved termite under sterile conditions in a laboratory where previous work on insect DNA had not been done. A number of control PCR amplifications were done to exclude any high-molecular mass DNA from the analysis, because such DNA would probably be of recent origin. In spite of these precautions, traces of several types of insect DNA sequences were recovered from the fossil material, which were 'dipteran in general and drosophilid in particular'. The most highly degraded DNA (<250 base pairs) contained some termite-like sequences as deduced from analysis of mitochondrial and nuclear ribosomal DNA; these were somewhat arbitrarily assigned as being representative of the fossil, whereas the other insect DNAs were deemed to be contaminants. It will now be important to assess the reproducibility of such findings. Amber preservation appears to provide a uniquely advantageous way of retaining ancient DNA sequences, because the DNA is largely dehydrated, partly protected from atmospheric oxygen, and not exposed to microbial contamination. It is disappointing in this context that no verified recoveries of DNA fragments have been made from less ancient (10^5 – 10^6 years old) but similarly entombed dehydrated fossils before attempts to isolate DNA 10^7 – 10^8 years old. In contrast to the careful study of the termite in amber, a widely publicized report on the recovery of 17–20-million-year-old DNA from an ancient magnolia leaf⁶⁸ exhibits serious deficiencies in that stringent PCR controls were not done. The leaf had been recovered from a wet deposit, and it was observed that high-molecular mass chloroplast DNA was recovered after a small number of PCR cycles. A reinvestigation of magnolia leaf material from the same *Clarkia* deposit, done elsewhere in a laboratory not working on chloroplast DNA, indicated that no plant DNA sequences could be recovered by PCR, although some (probably recent) high-molecular mass bacterial DNA was present⁶⁹. The apparent observation that fully hydrated plant DNA might be retained in high-molecular mass form for 20 million years is incompatible with the known properties of the chemical structure of DNA; it is, therefore, as incredible as the report on bacterial growth at 250 °C. Much better experimental documentation (preferably provided by meticulous nucleic acid biochemists) is required before such claims can be seriously considered.

Mutagenesis, carcinogenesis and ageing

Because DNA repair processes are not completely effective, the decay of the covalent structure of DNA may be expected to contribute in a major way to spontaneous mutagenesis⁷⁰ and by inference to carcinogenesis. Investigations of spontaneous mutation patterns in *E. coli*⁷¹ and mammalian cells indicate that both misreading of the DNA template during normal replication, and the occasional occurrence of damaged nucleotides in the template, are contributing factors. Several mechanisms for generation of spontaneous DNA lesions may be involved. Oxidative damage would appear to make only a minor contribution under most conditions: the experimentally determined oxygen- and hydrogen peroxide-derived mutational spectra in the human *hprt* gene are quite different from that of spontaneous mutagenesis⁷², and a DNA repair-deficient yeast strain exhibited

similarly elevated spontaneous mutation frequencies during aerobic and anaerobic growth⁷³. In addition to miscoding base derivatives, several types of primarily cytotoxic DNA lesions can generate occasional single-site mutations; these alterations include apurinic sites⁷, thymine glycol⁷⁴ and 3-methyladenine residues⁷⁵. The accumulated effect of replication and transcription of such damaged DNA templates has been proposed to be an important contributory factor in spontaneous carcinogenesis^{50,76}.

During evolution, cells acquired efficient DNA repair processes to permit very rapid correction of the predominant forms of spontaneous DNA damage. Water, oxygen and SAM resemble all other so-called 'group-specific agents' in that, in addition to the major ones, several distinct minor DNA lesions are generated. Some of these may only become relevant after slow accumulation, and for this reason may not be well repaired. One case in point may be deaminated 5-methylcytosine, which occurs less frequently but is more troublesome to cells than deaminated cytosine. It seems likely that certain very minor DNA lesions may have unexpectedly large biological effects due to their inefficient removal by DNA repair. It has often been speculated that cellular ageing may be a consequence of the gradual deterioration of the covalent structure of DNA. If this were correct, one might expect to find an accumulation of one or several unusual DNA base derivatives in very long-lived human cells such as neurons. Note that it is difficult to propose reasonable chemical models in this regard. Such a hypothetical DNA component must be chemically stable under physiological conditions, and this rules out many spontaneous lesions such as deoxyxanthosine or 7-methyldeoxyguanosine. Furthermore, the base derivative should be refractory to known versatile DNA repair processes. The complex nuclease that initiates nucleotide excision–repair of lesions causing major helix distortion also has some ability to act as a backup system in the repair of discrete lesions handled more effectively by other repair enzymes, for example apurinic/apyrimidinic sites, thymine glycol, and single-strand breaks with blocked 3'-termini⁷⁷. Few DNA lesions would escape survey by this activity. With regard to experimental studies, the ³²P postlabelling technique devised to detect damage in nonradioactive DNA⁷⁸ seems primarily suited to reveal bulky lesions which should be readily corrected by the nucleotide excision–repair pathway. Nevertheless, several compounds of unknown structures (designated 'I compounds') have been observed in nuclear DNA from aged tissue as well as in mitochondrial DNA by this method⁷⁸. Cells deficient in nucleotide excision–repair would be expected to be much more vulnerable to some forms of spontaneous DNA damage, because there is no known backup system to this excision–repair pathway. In this context, it is interesting that repair-defective cells from severe cases of xeroderma pigmentosum seem unable to remove a major oxygen radical-induced form of DNA damage, which might accumulate⁸⁴, because such patients suffer from a gradual but massive loss of neurons with accompanying mental deterioration over a period of decades and also exhibit a significantly increased frequency of internal cancers⁷⁹. Even in normal human cells, there would be little or no evolutionary reason to repair accumulating DNA lesions in stem cells or in neurons that might only become affected after 40–100 years. Chemically defined models for cellular ageing may be proposed along these lines, in addition to other hypotheses such as telomere shortening⁸⁰. For example, nonenzymatic DNA methylation by SAM generates the minor pyrimidine lesions 3-methylthymine and 3-methylcytosine. These are chemically stable, and active DNA repair of N3-methylated pyrimidines by DNA glycosylases has not been detected⁸¹. Unless they are removed by the nucleotide excision–repair pathway, such derivatives could accumulate in DNA *in vivo* over several decades and contribute to human ageing. □

T. Lindahl is in the Imperial Cancer Research Fund, Clare Hall Laboratories, South Mimms, Hertfordshire EN6 3LD, UK.

1. Ahern, T. J. & Klibanov, A. M. *Science* **228**, 1280–1284 (1985).
2. Kossiakoff, A. A. *Science* **240**, 191–194 (1988).
3. Lindahl, T. *J. biol. Chem.* **242**, 1970–1973 (1967).
4. Pace, N. R. *Cell* **65**, 531–533 (1991).
5. Lindahl, T. & Nyberg, B. *Biochemistry* **11**, 3610–3618 (1972).
6. Lindahl, T. & Karlström, O. *Biochemistry* **12**, 5151–5154 (1973).
7. Loeb, L. A. & Preston, B. D. *A. Rev. Genet.* **20**, 201–230 (1986).
8. Lindahl, T. & Andersson, A. *Biochemistry* **11**, 3618–3623 (1972).
9. Shapiro, R. in *Chromosome Damage and Repair* (eds Seeberg, E. & Kleppe, K.) 3–18 (Plenum, New York, 1981).
10. Doetsch, P. W. & Cunningham, R. P. *Mutat. Res.* **236**, 173–201 (1990).
11. Dianov, G., Price, A. & Lindahl, T. *Molec. cell. Biol.* **12**, 1605–1612 (1992).
12. Lindahl, T. & Nyberg, B. *Biochemistry* **13**, 3405–3410 (1974).
13. Lindahl, T. *Prog. Nucleic Acid Res. molec. Biol.* **22**, 135–192 (1979).
14. Ehrlich, M., Zhang, X.-Y. & Inamdar, N. M. *Mutat. Res.* **238**, 277–286 (1990).
15. Frederico, L. A., Kunkel, T. A. & Shaw, B. R. *Biochemistry* **29**, 2532–2537 (1990).
16. Krickner, M. C. & Drake, J. W. *J. Bact.* **172**, 3037–3039 (1990).
17. Duncan, B. K. & Miller, J. H. *Nature* **287**, 560–561 (1980).
18. Chen, J.-D. & Lacks, S. A. *J. Bact.* **173**, 283–290 (1991).
19. Impellizzeri, K. J., Anderson, B. & Burgers, P. M. J. *J. Bact.* **173**, 6807–6810 (1991).
20. Cooper, D. N. & Youssoufian, H. *Hum. Genet.* **78**, 151–155 (1988).
21. Rideout, W. M. III, Coetzee, G. A., Olumi, A. F. & Jones, P. A. *Science* **249**, 1288–1290 (1990).
22. Selker, E. U. A. *Rev. Genet.* **24**, 579–613 (1990).
23. Shen, J.-C., Rideout, W. M. III, & Jones, P. A. *Cell* **71**, 1073–1080 (1992).
24. Perutz, M. F. *J. molec. Biol.* **213**, 203–206 (1990).
25. Karran, P. & Lindahl, T. *Biochemistry* **19**, 6005–6011 (1980).
26. Dianov, G. & Lindahl, T. *Nucleic Acids Res.* **19**, 3829–3833 (1991).
27. Hill-Perkins, M., Jones, M. D. & Karran, P. *Mutat. Res.* **162**, 153–163 (1986).
28. Wink, D. A. *et al. Science* **254**, 1001–1003 (1991).
29. Dimple, B. A. *Rev. Genet.* **25**, 315–337 (1991).
30. Kasai, H. & Nishimura, S. *Nucleic Acids Res.* **12**, 2137–2145 (1984).
31. Shibutani, S., Takeshita, M. & Grollman, A. *Nature* **349**, 431–434 (1991).
32. Tchou, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 4690–4694 (1991).
33. Aruoma, O. I., Halliwell, B. & Dizdaroğlu, M. *J. biol. Chem.* **264**, 13024–13028 (1989).
34. Michaels, M. L., Cruz, C., Grollman, A. P. & Miller, J. H. *Proc. natn. Acad. Sci. U.S.A.* **89**, 7022–7025 (1992).
35. Maki, H. & Sekiguchi, M. *Nature* **355**, 273–275 (1992).
36. Fraga, C. G., Shigenaga, M. K., Park, J.-W., Degan, P. & Ames, B. N. *Proc. natn. Acad. Sci. U.S.A.* **87**, 4533–4537 (1990).
37. Wallace, S. S. *Envir. molec. Mutag.* **12**, 431–477 (1988).
38. Breimer, L. H. *Molec. Carcinog.* **3**, 188–197 (1990).
39. Kuo, C.-F. *et al. Science* **258**, 434–440 (1992).
40. Wagner, J. R., Hu, C.-C. & Ames, B. N. *Proc. natn. Acad. Sci. U.S.A.* **89**, 3380–3384 (1992).
41. Dirksen, M. L., Blakely, W. F., Holwitt, E. & Dizdaroğlu, M. *Int. J. Radiat. Biol.* **54**, 195–204 (1988).
42. Carmichael, P. L., Shē, M. N. & Phillips, D. H. *Carcinogenesis* **13**, 1127–1135 (1992).
43. Szatrowski, T. P. & Nathan, C. F. *Cancer Res.* **51**, 794–798 (1991).
44. Muret, J. F., Odin, F., Polverelli, M. & Cadet, J. *Chem. Res. Toxicol.* **3**, 102–110 (1990).
45. Joenje, H. *Mutat. Res.* **219**, 193–208 (1989).
46. Ljungman, M. & Hanawalt, P. C. *Molec. Carcinog.* **5**, 264–269 (1992).
47. Richter, C., Park, J.-W. & Ames, B. N. *Proc. natn. Acad. Sci. U.S.A.* **85**, 6465–6467 (1988).
48. Wallace, D. C. *et al. Science* **242**, 1427–1430 (1988).
49. Linnane, A. W., Marzuki, S., Ozawa, T. & Tanaka, M. *Lancet* **i**, 642–645 (1989).
50. Lutz, W. K. *Mutat. Res.* **238**, 287–295 (1990).
51. Paik, W. K., Lee, H. W. & Kim, S. *FEBS Lett.* **58**, 39–42 (1975).
52. Rydberg, B. & Lindahl, T. *EMBO J.* **1**, 211–216 (1982).
53. Barrows, L. R. & Magee, P. N. *Carcinogenesis* **3**, 349–351 (1982).
54. Park, J.-W. & Ames, B. N. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7467–7470 (1988).
55. Rebeck, G. W. & Samson, L. J. *Bact.* **173**, 2068–2076 (1991).
56. Vaughan, P., Sedgwick, B., Hall, J., Gannon, J. & Lindahl, T. *Carcinogenesis* **12**, 263–268 (1991).
57. Kodama, M. & Saito, H. *Cancer Lett.* **10**, 319–324 (1980).
58. Bucala, R. P., Model, P. & Cerami, A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 105–109 (1984).
59. Lee, A. T. & Cerami, A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8311–8314 (1987).
60. Setlow, P. *J. Bact.* **174**, 2737–2741 (1992).
61. Baross, J. A. & Deming, J. W. *Nature* **303**, 423–426 (1983).
62. Trent, J. D., Chastain, R. A. & Yayanos, A. A. *Nature* **307**, 737–740 (1984).
63. Higuchi, R., Bowman, R., Friedberger, M., Ryder, O. A. & Wilson, A. C. *Nature* **312**, 282–284 (1984).
64. Pääbo, S. *Nature* **314**, 644–645 (1985).
65. Pääbo, S., Higuchi, R. G. & Wilson, A. C. *J. biol. Chem.* **264**, 9709–9712 (1989).
66. Hagelberg, E. & Clegg, J. B. *Proc. R. Soc. B* **244**, 45–50 (1991).
67. De Sallé, R., Gatesy, J., Wheeler, W. & Grimaldi, D. *Science* **257**, 1933–1936 (1992).
68. Golenberg, E. M. *et al. Nature* **344**, 656–658 (1990).
69. Sidow, A., Wilson, A. C. & Pääbo, S. *Phil. Trans. R. Soc. B* **333**, 429–433 (1991).
70. Smith, K. C. *Mutat. Res.* **277**, 139–162 (1992).
71. Schaaper, R. M. & Dunn, R. L. *Genetics* **129**, 317–326 (1991).
72. Oller, A. R. & Thilly, W. G. *J. molec. Biol.* **228**, 813–826 (1992).
73. Ramotar, D., Popoff, S. C., Gralla, E. B. & Dimple, B. *Molec. cell. Biol.* **11**, 4537–4544 (1991).
74. Basu, A. K., Loechler, E. L., Leadon, S. A. & Essigmann, J. M. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7677–7681 (1989).
75. Klungland, A., Fairbairn, L., Watson, A. J., Margison, G. P. & Seeberg, E. *EMBO J.* **11**, 4439–4444 (1992).
76. Loeb, L. A. *Cancer Res.* **49**, 5489–5496 (1989).
77. Snowden, A., Kow, Y. W. & van Houten, B. *Biochemistry* **29**, 7251–7259 (1990).
78. Randerath, K., Leih, J. G., Gladek, A. & Randerath, E. *Mutat. Res.* **219**, 121–133 (1989).
79. Cleaver, J. E. & Kraemer, K. H. in *The Metabolic Basis of Inherited Disease* 6th edn. (eds Scriver, C. R. *et al.*) 2949–2971 (McGraw-Hill, New York, 1989).
80. Harley, C. B., Fitcher, A. B. & Greider, C. W. *Nature* **345**, 458–460 (1990).
81. McCarthy, T. V., Karran, P. & Lindahl, T. *EMBO J.* **3**, 545–550 (1984).
82. Lindahl, T. *Nature* **259**, 64–66 (1976).
83. Saporito, S. M., Geden, M. & Cunningham, R. P. *J. Bact.* **171**, 2542–2546 (1989).
84. Satoh, M. S., Jones, C. J., Wood, R. D. & Lindahl, T. *Proc. natn. Acad. Sci. U.S.A.* (in the press).

Young formation age of a mantle plume source

Cornelia Class, Steven L. Goldstein, Stephen J. G. Galer & Dominique Weis*

Max-Planck-Institut für Chemie, Postfach 3060, W-6500 Mainz, Germany

* Université Libre de Bruxelles, 50 Avenue F. D. Roosevelt, B-1050 Brussels, Belgium

The Ninetyeast Ridge hotspot track displays strontium, neodymium and lead isotope variations over time that reflect simple radioactive decay in the plume source rather than a change in the mantle components present. The lead isotope variations indicate that the time spent by the plume source in a non-convecting mantle boundary layer was only a few hundreds of millions of years, contrary to the conventional view of individual plume sources as old (1–3 Gyr) or persistent features.

MANTLE plumes are the main mode by which material upwells from the deep mantle^{1–3}. When the lithosphere moves over a plume, a hotspot track is formed, which is manifested as linear chains of ocean-island volcanoes, seamounts or aseismic ridges, and continental flood basalts^{1,4,5}. Such tracks provide the only means available of examining the compositional evolution of a plume through time, stemming from a common source region deep in the mantle. This record extends back to ~200 Myr in some cases⁶. Distances along each track usually correspond to past loci and ages of the plume. Within individual tracks, changes in initial (age-corrected) isotope ratios over time have been interpreted, conventionally, as reflecting changes in the ‘mantle components’^{7,8} present in the source (compare with refs

9–11). Here we examine the Ninetyeast Ridge plume track in some detail. Initial Nd and Pb isotope ratios of the plume source show large increases with decreasing age. Strontium isotope ratios increase slightly by an amount which is insignificant compared with the range of Sr isotope ratios in ocean basalts. These isotope characteristics would be unlikely to arise if a change in the mixture of ‘mantle components’ had taken place in the plume source with time. Rather, they seem primarily to reflect radioactive decay with time in the source, which we term an ‘evolving plume source’. The changes in isotope ratios limit the timing of chemical fractionations involved; they indicate that the Ninetyeast Ridge plume source is a young, ephemeral feature persisting for <1 Gyr. If these observations are generally

applicable to plume sources, they have important implications for the origin and longevity of these sources, as well as the rates of convective circulation in the mantle.

Evolving plume sources

If the isotope signature of a mantle plume evolves solely in response to radioactive decay in its source, which was isotopically homogeneous at the time of formation, initial isotopic compositions in the erupted basalts should evolve in a simple manner through time. This should reflect the radioactive half-lives of the long-lived nuclides in question, the parent-daughter ratios in the source, and the time available. If the basalts reflect the average composition of the plume source, the daughter isotope ratios should show systematic increases with time. If the parent to daughter ratios are low, the isotope ratios should remain nearly constant. This would be the case for $^{207}\text{Pb}/^{204}\text{Pb}$ ratios as a consequence of the low present-day $^{235}\text{U}/^{204}\text{Pb}$ ratios in the Earth and the limited time record (<200 Myr) of plume tracks. Initial $^{87}\text{Sr}/^{86}\text{Sr}$ ratios would also be expected to remain nearly constant over this time frame, as most mantle sources have low Rb/Sr ratios.

A plume track reflecting changes in source composition (that is, in the mantle components present^{7,8}) is unlikely to show the features of an evolving plume source. The magmatic processes that have been responsible for production of continental and oceanic crust over Earth history fractionate Rb/Sr ratios in the opposite sense to Sm/Nd ratios. After sufficient time for radioactive decay, higher $^{143}\text{Nd}/^{144}\text{Nd}$ ratios are accompanied by lower $^{87}\text{Sr}/^{86}\text{Sr}$ ratios. Such an anti-correlation is observed for ocean basalts as a whole, forming the basis for the Nd-Sr 'mantle

array'^{7,12}. This is also to be expected if the mixture of mantle components^{7,8} changed in the source: nearly every change would increase some isotope ratios while others decrease. In the terminology of Zindler and Hart⁸, decreasing contributions of a DMM (depleted mid-ocean-ridge basalt, or MORB, mantle) component would cause $^{87}\text{Sr}/^{86}\text{Sr}$ ratios to increase and $^{143}\text{Nd}/^{144}\text{Nd}$ ratios to decrease. Decreasing contributions of 'HIMU' or EM2 (high- μ and enriched-mantle 2 mantle components of ref. 8) would cause a decrease in Pb or Sr isotope ratios respectively. $^{87}\text{Sr}/^{86}\text{Sr}$ ratios vary over a large range in ocean basalts, from <0.703 to >0.708, and are particularly sensitive to a change in the mixture of mantle components. If the initial isotope ratios of other decay systems are increasing along a plume track while $^{87}\text{Sr}/^{86}\text{Sr}$ and $^{207}\text{Pb}/^{204}\text{Pb}$ ratios remain nearly constant, this is unlikely to be a consequence of changing mantle components in the plume source.

Isotope arrays and plume source compositions

When evaluating long-lived hotspot systems, practical considerations that must be taken into account include mixing unrelated to the plume and alteration of the older lavas. All continental flood-basalt provinces and many ocean-island suites show ranges in Sr, Nd and Pb isotope ratios. These isotopic arrays reflect either the mixing of isotopically distinct deep mantle sources or lithospheric contamination of the magmas¹³⁻¹⁵. Mixing, entrainment and contamination can be complex in plumes, and particularly so in the plume head during the initial phase¹⁶. Measurements on a sufficient number of representative samples of one age are required to delimit these isotopic arrays before the plume source can be distinguished from the exotic (extra-plume) material. In practice, the ability to follow a plume composition with time establishes criteria for distinguishing the 'true' plume source from these exotic components when evaluating isotopic arrays.

Accurate initial isotope ratios for the plume source in the past also have to be determined. This requires unravelling the radioactive decay within the source from the *in situ* radioactive decay in the lavas since their eruption. In older basalts, estimation of reliable initial $^{87}\text{Sr}/^{86}\text{Sr}$ ratios is difficult as both the $^{87}\text{Sr}/^{86}\text{Sr}$ and Rb/Sr ratios may have been modified by post-eruption alteration. We have used the lowest measured $^{87}\text{Sr}/^{86}\text{Sr}$ ratios on leached samples rather than the age-corrected $^{87}\text{Sr}/^{86}\text{Sr}$ ratios. Determining initial Pb isotope ratios is also problematic because the U/Pb ratio may similarly be affected. Where U and Pb concentrations are available, we have used the age-corrected ratios. We calculated initial $^{143}\text{Nd}/^{144}\text{Nd}$ ratios by assuming that the Sm/Nd and $^{143}\text{Nd}/^{144}\text{Nd}$ ratios are not significantly changed by alteration. We stress, however, that the trends discussed below do not depend on our preferred manner of age-correcting the data.

Ninetyeast Ridge hotspot track

The Ninetyeast Ridge plume (also called the Kerguelen-Heard plume) is thought to be responsible for the Rajmahal traps (northeast India) and the Kerguelen Plateau at ~117 Myr, followed later by the Ninetyeast Ridge, and Kerguelen and Heard Islands (compare ref. 6) (Fig. 1). The isotope data include cores drilled by DSDP legs 22 and 26 and ODP leg 121 which together recovered ~80-40-Myr-old basaltic basement from seven sites along ~5000 km of the Ninetyeast Ridge. The complete dataset for the track thus spans ~117 Myr.

The most striking aspect of the Ninetyeast Ridge plume track data is that the initial $^{206}\text{Pb}/^{204}\text{Pb}$ and $^{208}\text{Pb}/^{204}\text{Pb}$ ratios show a systematic increase with decreasing age while $^{207}\text{Pb}/^{204}\text{Pb}$ ratios remain nearly constant (Fig. 2). This extraordinary relationship is fully consistent with an evolving plume source. In detail, at some of the localities, the plume source composition is partly obscured by the presence of components extraneous to the plume as can be seen at Heard Island, the Rajmahal traps and the Kerguelen Plateau (Fig. 2). Mixing between these

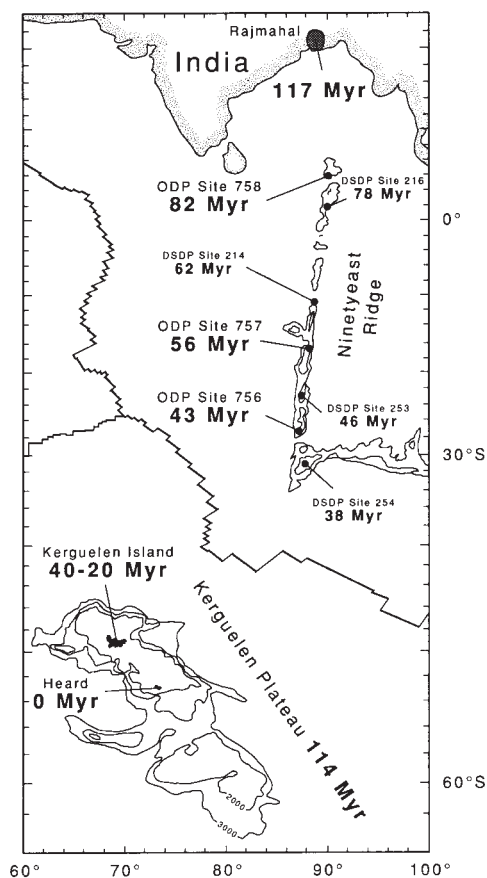


FIG. 1 Location map of the East Indian Ocean showing the Rajmahal traps, the Ninetyeast Ridge with drill sites from DSDP leg 22, 26 and ODP leg 121, Kerguelen Plateau, Kerguelen Archipelago and Heard Island, as well as the Indian mid-ocean-ridge axes. Contour interval is 1,000 m.

TABLE 1 Ninetyeast Ridge plume source chemistry

Average isotope composition of Ninetyeast Ridge plume source at various times*							Calculations of parent–daughter ratios†		
Age (Myr)	Location	²⁰⁶ Pb/ ²⁰⁴ Pb	²⁰⁷ Pb/ ²⁰⁴ Pb	²⁰⁸ Pb/ ²⁰⁴ Pb	¹⁴³ Nd/ ¹⁴⁴ Nd	⁸⁷ Sr/ ⁸⁶ Sr	Time interval (Myr)	μ	κ
117	Rajmahal	—	—	—	0.51263	0.70419	43–82	34	2.0
114	Kerguelen Plateau	18.164	15.552	38.522	0.51258	0.70417	56–82	39	1.7
82	Site 758	18.409	15.572	38.763	—	—	0–43	35	4.2
58	Site 757	18.568	15.592	38.846	—	—	0–56	32	4.0
43	Site 756	18.617	15.566	38.892	—	—	0–82	34	3.1
0	Heard	18.850	15.583	39.200	0.51278	0.7046	0–114	38	3.1

* Averages of the high initial $^{206}\text{Pb}/^{204}\text{Pb}$ and $^{208}\text{Pb}/^{204}\text{Pb}$ end of the arrays of the three ODP leg 121 sites and Kerguelen Plateau data were taken as the best estimate of the plume source composition at these various times. In the case of the Kerguelen Plateau, this coincides with the high- $^{143}\text{Nd}/^{144}\text{Nd}$ and low- $^{87}\text{Sr}/^{86}\text{Sr}$ end of the array formed by the data. Similarly, the average of the high- $^{143}\text{Nd}/^{144}\text{Nd}$ and low- $^{87}\text{Sr}/^{86}\text{Sr}$ end of the Rajmahal data array was taken as the plume source composition 117 Myr ago. The composition of the present-day plume source was assumed to be the endmember of the high- $^{206}\text{Pb}/^{204}\text{Pb}$ and $^{143}\text{Nd}/^{144}\text{Nd}$ end of the Heard Island mixing array. Data sources: Rajmahal¹⁴, Kerguelen Plateau^{13,57}, Ninetyeast Ridge ODP sites 756, 757 and 758¹⁰ and Heard Island¹⁵.

† μ -values ($^{238}\text{U}/^{204}\text{Pb}$) and κ -values ($^{232}\text{Th}/^{238}\text{U}$) are calculated from the difference in isotopic composition over the stated time interval.

components and the plume-derived material forms large isotopic arrays, and the composition of the plume source at each point in time in these cases is best represented by the highest $^{206}\text{Pb}/^{204}\text{Pb}$ and $^{143}\text{Nd}/^{144}\text{Nd}$ ratios in these arrays. If these mixing effects are unravelled, then the initial $^{143}\text{Nd}/^{144}\text{Nd}$ ratios also show increases with decreasing age while initial $^{87}\text{Sr}/^{86}\text{Sr}$ ratios remain between 0.7040 and 0.7045 (Fig. 3). Below, we discuss these trends in detail. We show that the data set, taken as a whole, provides a means of identifying the plume source composition within each sample location.

The secular isotopic changes in the plume source are obvious from the systematic increases in initial $^{206}\text{Pb}/^{204}\text{Pb}$ and $^{208}\text{Pb}/^{204}\text{Pb}$ ratios with time, as best defined by the samples from the Ninetyeast Ridge (Fig. 2). The situation is more complicated

for samples from other individual locations where the Pb isotopes often show larger variations. But evaluation of these Pb isotope arrays shows them to be mixtures of the same plume source plus an isotopically distinct component in each case. For both the Kerguelen Plateau and Heard Island samples, one end of the Pb isotope arrays (having the highest $^{206}\text{Pb}/^{204}\text{Pb}$ and $^{208}\text{Pb}/^{204}\text{Pb}$ ratios) approaches the value for the plume source predicted by the secular trend of Pb isotopes along the Ninetyeast Ridge (Fig. 2). Lead isotope ratios of Kerguelen do not form simple arrays (Fig. 2c), and the scatter makes it difficult to determine the plume source composition unambiguously. But with the exception of MORB-contaminated lavas^{17,18} erupted at ~40 Myr, when the plume was coincident with the Southeast Indian Ridge^{19,20}, the Pb isotope data all lie on the Ninetyeast

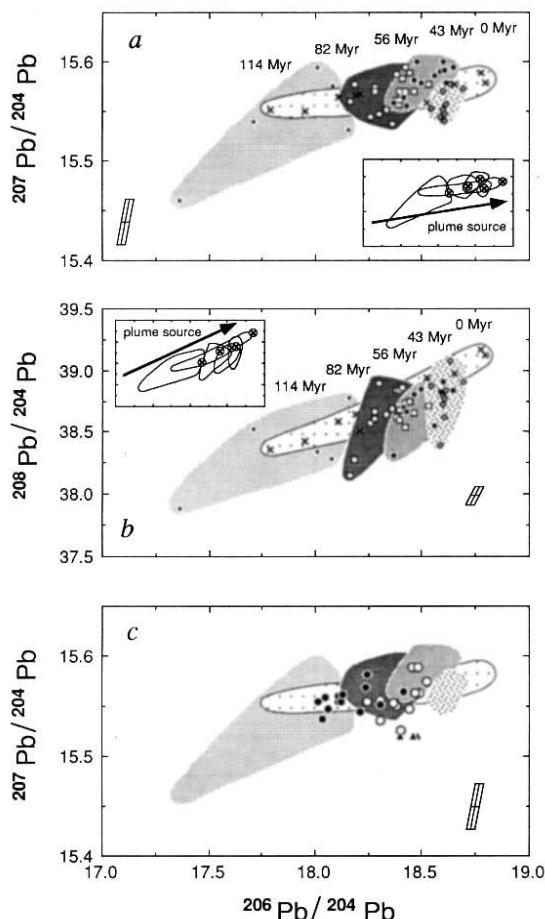


FIG. 2 a, b, Initial $^{207}\text{Pb}/^{204}\text{Pb}$ and $^{208}\text{Pb}/^{204}\text{Pb}$ ratios plotted against $^{206}\text{Pb}/^{204}\text{Pb}$ ratio for the Ninetyeast Ridge hotspot track. ♦: Kerguelen Plateau (~114 Myr; no initial $^{208}\text{Pb}/^{204}\text{Pb}$ available); □: ODP sites 758 (~82 Myr); ■: 757 (~56 Myr); ◇: 756 (~43 Myr); ×: Heard Island (0 Myr). Shaded fields delimit samples of the same age. Insets show best estimates of the plume source composition at various times (indicated by ⊗; see Table 1), as used to calculate μ and κ -values of the source (Table 1). In the insets, arrows illustrate the secular evolution expected from radioactive decay in a plume source of nearly constant composition. c, Same data fields as in a showing Kerguelen Archipelago samples: tholeiites ~40 Myr (▲; W. M. White, unpublished data), alkali basalts ~20 Myr (○), late stage alkaline rocks ~10 Myr (●). Data sources are given in Table 1, Kerguelen Islands data from ref. 18 and D.W. *et al.* (manuscript in preparation). Error bars representing typical analytical uncertainties for individual datapoints are shown for each plot.

Ridge secular trend. (Although these lavas have K-Ar dates of ~26 Myr (ref. 21), Giret *et al.*²² have argued that these ages probably reflect thermal resetting and that the eruption age is ~40 Myr.) Therefore, the changes in the Pb isotopes over the history of the plume track are consistent with derivation from an evolving plume source.

The Heard Island, Kerguelen and Kerguelen Plateau samples also display large variations in Nd and Sr isotope ratios. The highest $^{206}\text{Pb}/^{204}\text{Pb}$ ratios are associated with the highest Nd and lowest Sr isotope ratios (Fig. 3). In the Kerguelen Plateau, these Sr and Nd isotope ratios are in the range expected for the plume, based on the samples from the Ninetyeast Ridge. For Heard Island, the predicted values for the present day lie directly on the extension of its Nd-Sr isotope array (Fig. 3). The Rajmahal traps, also associated with the Ninetyeast Ridge plume, show a large range in Sr and Nd isotope ratios (no Pb isotope data are available), which can be understood in a manner similar to the Kerguelen Plateau (Fig. 3). In the Ninetyeast Ridge itself, most samples have nearly the same $^{87}\text{Sr}/^{86}\text{Sr}$ ratios as the inferred plume source compositions at the initial stage of plume activity, and $^{143}\text{Nd}/^{144}\text{Nd}$ ratios intermediate between that and the present day. An exception is the 43-Myr-old ODP site 756, whose samples have high Nd and low Sr isotope ratios, plotting in a distinct field away from the other Ninetyeast Ridge data (Fig. 3). When these lavas erupted, the position of the Ninetyeast Ridge plume coincided with the Southeast Indian Ridge²⁰, and the Nd and Sr isotope ratios clearly record contamination by shallow MORB mantle¹⁰. The Pb isotope ratios have not been as strongly affected although $^{207}\text{Pb}/^{204}\text{Pb}$ ratios appear to be slightly low (Fig. 2), consistent with a MORB contribution. The two analysed samples from DSDP site 216 have high Sr and low Nd isotope ratios relative to the Ninetyeast Ridge plume source, lying in the middle of the present-day Heard Island mixing array. Either they are an exception to the plume source trend displayed by the other sites, or they lie on a mixing array and do not reflect the plume source. To distinguish between these possibilities, more measurements from this site are needed.

In summary, we have shown that there are regular increases in initial $^{206}\text{Pb}/^{204}\text{Pb}$, $^{208}\text{Pb}/^{204}\text{Pb}$ and $^{143}\text{Nd}/^{144}\text{Nd}$ ratios along the Ninetyeast Ridge hotspot track through time, no detectable changes in $^{207}\text{Pb}/^{204}\text{Pb}$, and small changes in $^{87}\text{Sr}/^{86}\text{Sr}$ ratios. It is unlikely that these shifts reflect changes in the mixture of mantle components in the source. For this to be the case, over the 117-Myr history of the plume, the different mixtures of mantle components must have had, coincidentally, the same $^{207}\text{Pb}/^{204}\text{Pb}$ ratios and nearly the same $^{87}\text{Sr}/^{86}\text{Sr}$ ratios, while the other isotope ratios varied. A more likely explanation is that the isotopic evolution of the Ninetyeast Ridge plume reflects an evolving plume source in which secular changes in isotope ratios are solely the result of radioactive decay. We have argued

elsewhere that these features are shared by other hotspot tracks²³. These include the Paraná-Etendeka-Walvis Ridge-Tristan, Karoo-Madagascar-Marion, North Atlantic Tertiary Province-Iceland and Ontong Java-Louisville hotspot tracks. Thus, the effects of radioactive decay are clearly discernible and seem to be primarily responsible for the changes in isotope ratios with time seen along many hotspot tracks.

Plume source composition

The regular secular increases in isotope ratios exhibited along the Ninetyeast Ridge allow us to infer aspects of the plume source chemistry. In the simplest case, these changes in initial isotope ratios through time (Figs 2 and 3) directly reflect the average parent-daughter ratios of the plume source in a boundary layer in the mantle. To calculate these parent-daughter ratios, our best estimate of the plume source composition uses the highest initial $^{206}\text{Pb}/^{204}\text{Pb}$ and $^{208}\text{Pb}/^{204}\text{Pb}$ isotope ratios of the arrays at each age, to exclude the effects of mixing with extra-plume components. The $^{206}\text{Pb}/^{204}\text{Pb}$ ratio of the present-day plume is assumed to be the same as one mixing endmember already defined for Heard Island¹⁵. The values of μ ($=^{238}\text{U}/^{204}\text{Pb}$) and κ ($=^{232}\text{Th}/^{238}\text{U}$) were calculated using the estimated plume source compositions for Heard Island, the intermediate-age ODP sites, and the Kerguelen Plateau in different combinations (Table 1). The total ranges are $\mu = 32\text{--}39$, giving a best estimate of 35, and $\kappa = 1.7\text{--}4.2$ (Table 1). Using average $^{206}\text{Pb}/^{204}\text{Pb}$ ratios for the intermediate age ODP sites instead of the highest ratios makes little difference to the results. The $^{87}\text{Rb}/^{86}\text{Sr}$ and $^{147}\text{Sm}/^{144}\text{Nd}$ ratios are less well constrained because of the limited changes in $^{87}\text{Sr}/^{86}\text{Sr}$ and $^{143}\text{Nd}/^{144}\text{Nd}$ ratios over the history of the plume, and alteration effects on Rb/Sr and $^{87}\text{Sr}/^{86}\text{Sr}$ ratios. Our best estimates are based on the isotopic differences between the Kerguelen Plateau, Rajmahal traps and Heard Island, spanning the longest time interval, and yield $^{147}\text{Sm}/^{144}\text{Nd}$ of 0.23 ± 0.11 and $^{87}\text{Rb}/^{86}\text{Sr}$ of 0.25 ± 0.15 . The associated errors were estimated from the parent-daughter ratios calculated from the maximum and minimum values of Nd and Sr isotope ratios for the plume source over time, as indicated by the data in Fig. 3. These ratios are also consistent with estimates derived from the lavas erupted at intermediate ages.

The true significance of these estimated elemental ratios in the Ninetyeast Ridge hotspot source is an intriguing question. How well do they reflect the actual plume source ratios? The inferred κ , $^{147}\text{Sm}/^{144}\text{Nd}$, and $^{87}\text{Rb}/^{86}\text{Sr}$ ratios, especially when the large errors are taken into account, are plausible values for plume sources, and comparable to the ratios typically observed in the erupted lavas. Although some oceanic island basalts have μ values of ~35 (ref. 24), as inferred for the Ninetyeast Ridge hotspot source (Table 1), this value is higher than those found

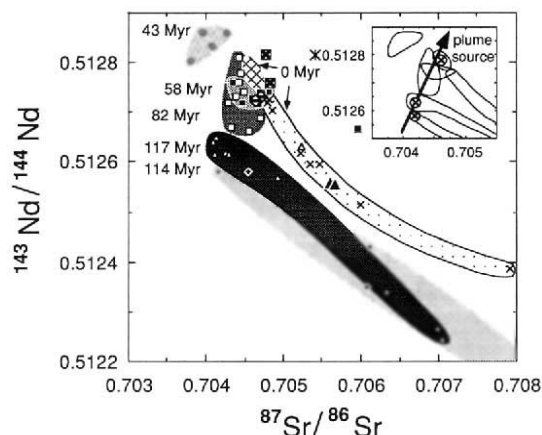


FIG. 3 Initial $^{143}\text{Nd}/^{144}\text{Nd}$ ratios plotted against initial or leached $^{87}\text{Sr}/^{86}\text{Sr}$ ratios for the Ninetyeast Ridge hotspot track. Symbols as in Fig. 2a, b; in addition, Δ : Rajmahal (~117 Myr); $\boxtimes$: DSDP sites 214 (~62 Myr); $\blacktriangle$: 216 (~77 Myr); $*$: 253 (~46 Myr); $\diamond$: 254 (~38 Myr). Cross-hatched field shows the extension of the Heard Island mixing array (see text for explanation). Inset shows the best estimate of the plume source composition ($\otimes$; see Table 1) used to calculate $^{87}\text{Rb}/^{86}\text{Sr}$ and $^{147}\text{Sm}/^{144}\text{Nd}$ values of the source (see text). The arrow marks the trend of the evolving plume source composition with time. Data sources are listed in Table 1, with additional Ninetyeast Ridge data from refs 14, 58.

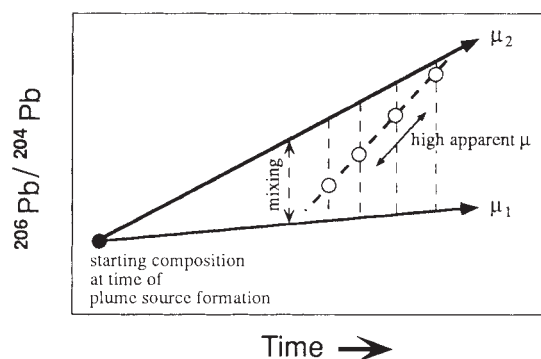


FIG. 4 $^{206}\text{Pb}/^{204}\text{Pb}$ ratio plotted against time showing how a high apparent μ can be produced by changing the mixture between two differently enriched ($\mu_2 > \mu_1$) regions in the plume source. In this example, the two regions have evolved independently since the time of plume source formation and have common initial isotope ratios. An increasing amount of high- μ_2 material in the mixture with time is necessary to produce the observed secular evolution along the Ninetyeast Ridge hotspot track (Fig. 1). This could explain the high apparent μ s observed, but is problematic for reasons discussed in the text.

in the least altered Ninetyeast Ridge lavas, for which $\mu \sim 10\text{--}30$. The U/Pb ratio in the source could be higher than in the lavas if Pb was more incompatible than U during partial melting and fractional crystallization^{25,26}, but the consensus view from partitioning experiments and other arguments suggests the reverse to be true^{27–30}. Therefore, if these parent-daughter ratios truly reflect the plume source, a problem exists as to how a high- μ source can generate low- μ basalts. It is conceivable that this might be due, for example, to a post-eruptive gain of Pb over U in the basalts.

Alternatively, it is possible that the μ inferred from the isotopic evolution is higher than the true source μ . A full discussion of this issue will be given elsewhere (C.C., S.L.G. and S.J.G.G., manuscript in preparation). We can illustrate it with one example in which a plume source was formed with uniform initial isotope ratios but where the U/Pb ratios are variable (Fig. 4). The $^{206}\text{Pb}/^{204}\text{Pb}$ ratios would increase at a slower rate in the low- μ regions of the plume source than in the high- μ regions. If magmas were derived increasingly from the higher- μ regions of the plume with time, the observed Pb isotope ratios would increase rapidly. The μ inferred from the Pb isotopic evolution would then be higher than either component in the plume source. This example would still be an evolving plume source, because the source formed with nearly uniform isotopic compositions and the secular changes in the isotope ratios are solely the result of radioactive decay following its formation (Fig. 4). The question then arises as to what circumstances would allow the appropriate coupled secular increase in both the Nd and Pb isotope ratios in the erupted basalts, with the increase in $^{206}\text{Pb}/^{204}\text{Pb}$ ratios implying a higher μ than is actually present in the plume source. In the case of the Ninetyeast Ridge plume, these conditions are met only if the Sm/Nd ratio is nearly the same in both the low- and high- μ components. Therefore, a special process is first required that enriches U/Pb ratios independently of Sm/Nd, and a second special process is required to sample the high- μ portions of the plume increasingly with time. Although a mechanism of this sort cannot be precluded, it is improbable, because it is unclear how such a combination of events could come about.

Life span of the hotspot source

The variations in Pb, Sr and Nd isotope ratios in oceanic basalts have been viewed as reflecting either the ages of ancient mantle differentiation events^{31–33}, or recent mixtures of components having distinctive isotope ratios and independent histories^{7,8,34}. In both cases, the mantle sources for plumes and ocean islands have been considered to require long periods of isolation (1–3 Gyr) from the convecting part of the mantle, probably within a mantle boundary layer, for the isotope heterogeneities to develop. Here we estimate the maximum total lifetime of the Ninetyeast Ridge plume source by considering what limitations can be placed on its present age, and how long it might continue to survive.

If the changes in the isotopic compositions seen in a plume track through time are due primarily to radioactive decay in the plume source, then the storage time of the plume source in the mantle can be estimated. One way is from the rate of growth of $^{206}\text{Pb}/^{204}\text{Pb}$ ratios, assuming that this directly reflects the μ -value of the plume source. An alternative is from the relative changes in $^{207}\text{Pb}/^{204}\text{Pb}$ and $^{206}\text{Pb}/^{204}\text{Pb}$ ratios through time, taking advantage of the coupled decay of the two isotopes of U to two of Pb. The latter method requires only that the variations in $^{207}\text{Pb}/^{204}\text{Pb}$ and $^{206}\text{Pb}/^{204}\text{Pb}$ ratios do not arise from mixing of components where one is from outside the evolving plume source.

If the μ of ~ 35 describes the isotopic evolution of the Ninetyeast Ridge plume track before the initial surface expression at ~ 117 Myr ago, fairly firm limits can be placed on the storage time of the plume source in the mantle. Extrapolated back in time, the $^{206}\text{Pb}/^{204}\text{Pb}$ ratio decreases markedly, accompanied by only a modest decrease in the $^{207}\text{Pb}/^{204}\text{Pb}$ ratio (Fig. 5a, b). The Pb isotopic composition of the plume source at present lies to the right of the present-day geochron, but only 300 Myr ago it would have been to the left of the geochron of that time (Fig. 5a). The Pb isotope ratios inferred for the plume source at 500 Myr ago ($^{206}\text{Pb}/^{204}\text{Pb} \approx 16.0$ and $^{207}\text{Pb}/^{204}\text{Pb} \approx 15.3$) fall close to growth curves for the evolution of Pb in the continents and mantle (Fig. 5b) and lie within the range typical of Proterozoic granulites³⁵. By 700 Myr ago, its Pb isotope composition would be irreconcilable with any reasonable Pb isotope evolution for terrestrial samples; such values are not even found in granulite facies rocks, which often have low μ and correspondingly low $^{206}\text{Pb}/^{204}\text{Pb}$ and high $^{207}\text{Pb}/^{204}\text{Pb}$ present-day ratios³⁵.

Thus, if the rate of increase of Pb isotope ratios seen over the history of the plume track extends to the period prior to its surface expression, its isolation time from the convecting mantle is certainly less than 700 Myr and more likely to be only a few hundred million years. Assuming that plumes rise from a boundary layer in the mantle, either the effect generating the high rate of $^{206}\text{Pb}/^{204}\text{Pb}$ increase occurred within the boundary layer, which we consider unlikely, or the amount of time spent there was no longer than a few hundred million years.

To address the question of how long the plume will survive, we use the relationship between Pb isotope ratios of ocean islands and MORB. All ocean islands measured, comprising about 80 oceanic islands from about 40 island groups, lie on or to the left of a Pb isotopic array formed by MORB and the HIMU islands⁸ (an exception is Guadalupe). This appears to be a robust feature of the sub-oceanic mantle. Unless this relationship is unrepresentative of the Earth throughout its history, we can assume that it held in the past and will hold into the future. We can therefore infer that the plume will probably disappear before its Pb isotopic composition falls to the right of a future MORB array. Initial Pb isotope ratios of 750-Myr-old ophiolites³⁶ form an array slightly to the left of the present-day MORB array, which will continue to evolve further

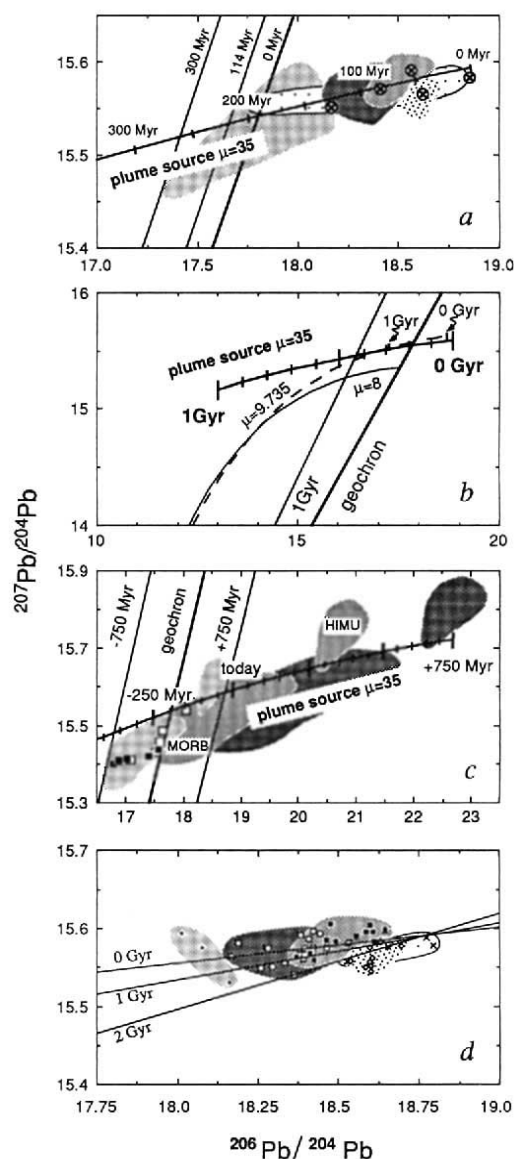
to the right of its present position. The slope of the MORB-HIMU array corresponds to a Pb-Pb isochron age of ~ 1.7 Gyr. The highest reasonable values for the Pb isotope ratios at ~ 750 Myr in the future would be described by an array corresponding to a ~ 2.5 -Gyr isochron. We assume that the least radiogenic values lie slightly to the left of the geochron of that time (like present-day MORB). A maximum estimate of the survival time would then be when the composition of the plume source composition is to the right of the array. Evolving at a rate defined by a μ of ~ 35 , the Ninetyeast Ridge plume source will develop Pb isotope ratios far to the right of this estimated MORB-HIMU array before ~ 750 Myr into the future (Fig. 5c). Therefore, an upper limit to the future lifetime of the plume is less than 750 Myr, and more likely just a few hundred million years. The present rate of magmatism is much smaller than that which formed the Ninetyeast Ridge originally (ref. 37, and N. H. Sleep, personal communication), suggesting that the plume may already be in its waning stage of activity.

Alternatively, the apparent μ -value inferred from the rate of increase in Pb isotope ratios might not reflect the true μ -value of the plume source (for example, as illustrated in Fig. 4). Even in this case, however, the age of the plume source can be constrained by the age-corrected $^{207}\text{Pb}/^{204}\text{Pb}$ and $^{206}\text{Pb}/^{204}\text{Pb}$ ratios over the history of the plume track. In a plume source

having a long residence time in a mantle boundary layer, low $^{206}\text{Pb}/^{204}\text{Pb}$ ratios would be accompanied by low $^{207}\text{Pb}/^{204}\text{Pb}$ ratios. This is not seen in samples from the Ninetyeast Ridge plume: $^{207}\text{Pb}/^{204}\text{Pb}$ ratios show little change despite a large variation in $^{206}\text{Pb}/^{204}\text{Pb}$ ratios. In Fig. 5d, the initial Pb isotope ratios of the Ninetyeast Ridge plume are compared to the variations expected if the plume source has ages of 0 to 2 Gyr. Although the initial isotope ratios of lavas with ages between 0 and ~ 120 Myr cannot be considered strictly to form an isochron, this comparison is justified as the coupled variations of $^{207}\text{Pb}/^{204}\text{Pb}$ and $^{206}\text{Pb}/^{204}\text{Pb}$ ratios are closely approximated by the isochrons. Our use of the Pb isochron slopes here is fundamentally different from that of ref. 32, where estimates of plume source ages were based on the Pb isotopic arrays from individual ocean islands; as discussed previously, these seem merely to reflect mixing of components extraneous to the plume¹⁵. The Pb isotopic compositions of the Ninetyeast Ridge plume source through time form an array which is shallower than the slopes of the 1-Gyr isochron (Fig. 5d), again constraining the age of the plume source to be < 1 Gyr.

In summary, the rate of increase of $^{206}\text{Pb}/^{204}\text{Pb}$ ratios over the history of Ninetyeast Ridge hotspot indicates that the maximum life span of the plume source is ~ 1.2 Gyr and more likely much less than 1 Gyr. This would encompass its creation

FIG. 5 $^{207}\text{Pb}/^{204}\text{Pb}$ ratios plotted against $^{206}\text{Pb}/^{204}\text{Pb}$ ratios showing the evolution of the Ninetyeast Ridge hotspot source composition through time (note the scale changes between diagrams). *a*, Data fields and best estimates of the plume source composition at the times of eruption ($\otimes$) shown. The evolution curve for the Ninetyeast Ridge plume source is calculated using $\mu = 35$ (Table 1). The present-day plume source composition is marked (0 Myr), together with its composition 100 and 300 Myr ago. The geochron today and the 114 and 300 Myr geochrons are shown as a reference frame. *b*, Backward calculation of the Ninetyeast Ridge hotspot source composition using $\mu = 35$ as in *a*, from present day to 1 Gyr ago. Also shown for reference are: the present-day geochron, the geochron 1 Gyr ago and the single-stage ($\mu = 8$) and Stacey-Kramers⁵⁹ (the wiggly arrows point to compositions at 1 Gyr ago and at 0 Gyr on the $\mu = 9.735$ dashed line) evolution curves for Pb, which represent the average mantle and probably upper crustal evolution, respectively. By > 500 Myr ago, the Pb isotope composition of the plume source would be irreconcilable with any reasonable Pb isotope evolution for terrestrial samples (see text). *c*, Forward calculation of the Ninetyeast Ridge hotspot source composition ($\mu = 35$) to 750 Myr in the future. The present day geochron, the geochron at 750 Myr before present (BP) and the future +750 Myr geochron are also shown. The shaded fields correspond to present-day data from mid-ocean-ridge basalts (MORB) and volcanic rocks from St Helena and Tubuaii Islands (HIMU). It is assumed that the MORB-HIMU data lie on a 1.7-Gyr isochron, and from this their past (750 Myr ago; light grey) and future (750 Myr from now; dark grey) positions have been calculated. ■, Initial Pb isotope ratios of whole rocks; □, feldspars, from ~ 750 -Myr-old ophiolite complexes (refs 36, 48). The Ninetyeast Ridge hotspot source composition will develop Pb isotope ratios to the right of this estimated MORB-HIMU array within ~ 750 Myr. *d*, $^{207}\text{Pb}/^{204}\text{Pb}$ - $^{206}\text{Pb}/^{204}\text{Pb}$ isochrons of 0, 1 and 2 Gyr are compared with the secular evolution in initial Pb isotopic compositions for the Ninetyeast Ridge hotspot (mixing arrays have been omitted), which form an array shallower than the slopes of the 1-Gyr isochron, constraining the age of the plume source to be < 1 Gyr.



as an entity, transport to a boundary layer in the mantle, thermal equilibration and upwelling, and its final disappearance. Independently, the coupled variation of $^{207}\text{Pb}/^{204}\text{Pb}$ and $^{206}\text{Pb}/^{204}\text{Pb}$ ratios also constrains the age of the source to be younger than ~ 1 Gyr. Such a period is far shorter than most geochemical estimates of likely storage times in a boundary layer in the mantle^{8,31–33,38,39}. It is, however, consistent with the results of numerical simulations of convective flow^{40–43} and Pb residence-time arguments^{44,45} which suggest that chemical heterogeneities are remixed into the convecting mantle within a few hundred million years. We have elsewhere argued that these observations are applicable to other hotspot tracks²³. Taking this into account, we conclude that the sources of many plumes seem to be relatively young and ephemeral features.

Plume source formation

The source characteristics inferred for the Ninetyeast Ridge hotspot strongly limit the choice of possible precursor materials and the processes involved in its formation. It has been suggested that plume sources are derived from subduction of oceanic crust and sediment³⁸, or delamination of continental lithosphere during either orogenesis⁴⁶ or rifting⁴⁷. These processes are claimed to be able to produce the spectrum of isotope variations found in ocean island basalts.

The precursor material that formed the Ninetyeast Ridge source was characterized by high $^{207}\text{Pb}/^{204}\text{Pb}$ ratios and comparatively low $^{206}\text{Pb}/^{204}\text{Pb}$ ratios. The suitability of old oceanic crust in this regard can be assessed using the isotopic compositions of ophiolites, some of which are as old as our inferred isolation time for the Ninetyeast Ridge hotspot source. The initial $^{206}\text{Pb}/^{204}\text{Pb}$ ratios of ~ 750 Myr-old ophiolites^{36,48} are far too high for them to serve as precursor material for the Ninetyeast Ridge source (Fig. 5c). Addition of sediments to the mantle along with oceanic crust is unlikely to change the bulk composition of the subducted material towards lower $^{206}\text{Pb}/^{204}\text{Pb}$ isotope ratios: all measured sediments on the ocean floor today have Pb isotope ratios lying to the right of the geochron⁴⁹. Neither existing ocean floor (with or without sediments) nor ophiolites seem to have the requisite Pb isotopic compositions. Therefore, provided that the rate of increase of Pb isotope ratios seen along the Ninetyeast Ridge plume track can be extended back to its earlier period in a mantle boundary

layer, former subducted oceanic crust seems to be ruled out as the origin for this source.

A more promising candidate for the precursor material, having low Pb and Nd isotope ratios, is recycled continental lithosphere. Granulite facies lower-crustal rocks often have low U/Pb ratios and, as a result, low $^{206}\text{Pb}/^{204}\text{Pb}$ and high $^{207}\text{Pb}/^{204}\text{Pb}$ ratios^{35,50}. This is exactly the isotopic signature deduced for the Ninetyeast Ridge plume source only a few hundred million years ago. Lower-crustal rocks also have Nd isotope ratios lower than the bulk Earth ratio and variable but often low Sr isotope ratios^{51,60}. This material therefore has the appropriate isotopic characteristics for the source. Some continental magmas indicate that such features also extend to regions of the lithospheric mantle⁵². We suggest, therefore, that the deep mantle source for this plume represents parts of delaminated lower crust or lithospheric mantle recycled back to the convecting mantle^{46,47,53–56}. Because the precursor must have low U/Pb ratios, an event is required which increases this ratio. As the plume source has a young formation age we suggest that the increase in U/Pb ratios is associated with the delamination process. One mechanism that can be excluded is extraction of Pb into the Earth's core: assuming the plume was derived from the core-mantle boundary, this would result in an additional deficit in siderophile elements which is not observed²⁸.

The continental lithosphere thus seems to be important in the generation of the gross isotopic variations in mantle-derived basalts. To the extent that the young age of the Ninetyeast Ridge plume source is applicable to other plume sources²³ and to MORB sources as well^{44,45}, the geochemical evidence favours rapid and efficient mixing in the convecting mantle on the large as well as the small scale. The total lifetime of many plume sources from initial formation to final disappearance seems to be ≤ 1 Gyr. Nevertheless, the large-scale variations of isotope ratios in mantle-derived rocks, such as those indicated by differences in $^{207}\text{Pb}/^{204}\text{Pb}$ ratios in ocean basalts, demand long isolation times and ancient (>1 Gyr) chemical fractionations. Such isotope heterogeneities can apparently only survive in regions isolated from the main mantle convective flow. If the residence times of material in mantle boundary layers are as short as we infer, the only viable candidate left for long-term storage within the silicate Earth is the continental lithosphere. □

Received 4 December 1992; accepted 19 March 1993.

- Morgan, W. J. *Nature* **230**, 42–43 (1971).
- Sleep, N. H. *J. geophys. Res.* **95**, 6715–6736 (1990).
- Davies, G. F. *Earth planet. Sci. Lett.* **99**, 94–109 (1990).
- Wilson, J. T. *Can. J. Phys.* **41**, 863–870 (1963).
- Morgan, W. J. in *The Sea, 7, The Oceanic Lithosphere* (ed. Emiliani, C.) (Wiley, New York, 1981).
- Duncan, R. A. & Richards, M. A. *Rev. Geophys.* **29**, 31–50 (1991).
- White, W. M. *Geology* **13**, 115–118 (1985).
- Zindler, A. & Hart, S. A. *Rev. Earth planet. Sci.* **14**, 493–571 (1986).
- Le Roex, A. P., Cliff, R. A. & Adair, J. I. *J. Petrol.* **31**, 779–812 (1990).
- Weis, D. & Frey, F. A. *Proc. ODP Sci. Res.* **121**, 591–610 (1991).
- Mahoney, J. J., Nicollet, C. & Dupuy, C. *Earth planet. Sci. Lett.* **104**, 350–363 (1991).
- O'Nions, R. K., Hamilton, P. J. & Evensen, N. M. *Earth planet. Sci. Lett.* **34**, 13–22 (1977).
- Weis, D., Bassias, Y., Gautier, I. & Mennessier, J.-P. *Geochim. cosmochim. Acta* **53**, 2125–2131 (1989).
- Mahoney, J. J., Macdougall, J. D., Lugmair, G. W. & Gopalan, K. *Nature* **303**, 385–389 (1983).
- Barringer, J. & Goldstein, S. L. *Nature* **348**, 59–62 (1990).
- Griffiths, R. W. & Campbell, I. H. *Earth planet. Sci. Lett.* **99**, 66–78 (1990).
- Storey, M. et al. *Nature* **336**, 371–374 (1988).
- Gautier, I., Giret, A., Loubet, M., Vidal, P. & Weis, D. *Earth planet. Sci. Lett.* **100**, 59–76 (1990).
- Schlich, R. *Mem. Hors Ser. Soc. Geol. Fr.* **6**, 103 (1975).
- Royer, J.-Y. & Sandwell, D. T. *J. geophys. Res.* **94**, 13,755–13,782 (1989).
- Watkins, N. D., Gunn, B. M., Nougier, J. & Baksi, A. K. *Bull. geol. Soc. Am.* **85**, 201–212 (1974).
- Giret, A., Cantagrel, J.-M. & Nougier, J. *C.R. Acad. Sci. Paris* **293**, Serie II, 191–194 (1990).
- Class, C., Goldstein, S. L., Galer, S. J. G. & Weis, D. V. M. *Goldschmidt Conf. Abstr.* 21 (Geochemical Society, 1992).
- Halliday, A. N. et al. *Nature* **359**, 623–627 (1992).
- Meijer, A., Kwon, T.-T. & Tilton, G. R. *J. geophys. Res.* **95**, 433–448 (1990).
- Qin, Z. *Earth planet. Sci. Lett.* **112**, 75–90 (1992).
- Tatsumoto, M. *Earth planet. Sci. Lett.* **38**, 63–87 (1978).
- Newsom, H. E., White, W. M., Jochum, K. P. & Hofmann, A. W. *Earth planet. Sci. Lett.* **80**, 299–313 (1986).
- Hofmann, A. W., Jochum, K. P., Seufert, M. & White, W. M. *Earth planet. Sci. Lett.* **79**, 33–45 (1986).
- Walker, D., Beattie, P. & Jones, J. H. *Eos* **73**, 616 (1992).

- Brooks, C., Hart, S. R., Hofmann, A. & James, D. E. *Earth planet. Sci. Lett.* **32**, 51–61 (1976).
- Chase, C. G. *Earth planet. Sci. Lett.* **52**, 277–284 (1981).
- Silver, P. G., Carlson, R. W. & Olson, P. A. *Rev. Earth planet. Sci.* **16**, 477–541 (1988).
- Zindler, A., Jagoutz, E. & Goldstein, S. L. *Nature* **298**, 519–523 (1982).
- Rudnick, R. L. & Goldstein, S. L. *Earth planet. Sci. Lett.* **98**, 192–207 (1990).
- Pallister, J. S., Stacey, J. S., Fischer, L. B. & Premo, W. R. *Precamb. Res.* **38**, 1–54 (1988).
- Sleep, N. H. *J. geophys. Res.* **97**, 20007–20019 (1992).
- Hofmann, A. W. & White, W. M. *Earth planet. Sci. Lett.* **57**, 421–436 (1982).
- Hart, S. R. *Nature* **309**, 753–757 (1984).
- Christensen, U. *Earth planet. Sci. Lett.* **95**, 382–394 (1989).
- Gurnis, M. & Davies, G. F. *J. geophys. Res.* **91**, 6375–6395 (1986).
- Gurnis, M. & Davies, G. F. *Geophys. Res. Lett.* **13**, 541–544 (1986).
- Hoffman, N. R. A. & McKenzie, D. P. *Geophys. J. R. astr. Soc.* **82**, 163–206 (1985).
- Galer, S. J. G. & O'Nions, R. K. *Nature* **316**, 778–782 (1985).
- Goldstein, S. & Chauvel, C. V. M. *Goldschmidt Conf. Abstr.* 48 (Geochemical Society, 1990).
- McKenzie, D. & O'Nions, R. K. *Nature* **301**, 229–231 (1983).
- Arndt, N. T. & Goldstein, S. L. *Tectonophysics* **161**, 201–212 (1989).
- Zimmer, M. thesis, Johannes Gutenberg-Universität Mainz (1989).
- Ben Othman, D., White, W. M. & Patchett, P. J. *Earth planet. Sci. Lett.* **94**, 1–21 (1989).
- Heier, K. S. *Phil. Trans. R. Soc. Lond. A* **273**, 429–442 (1973).
- Rudnick, R. L. in *Continental Lower Crust* (eds Fountain, D. M., Arculus, R. & Kay, R. W.) 269–316 (Elsevier, London, 1992).
- Dudas, F. O., Carlson, R. W. & Eggler, D. H. *Geology* **15**, 22–25 (1987).
- Molnar, P. & Gray, D. *Geology* **7**, 58–62 (1979).
- Houseman, G. A., McKenzie, D. P. & Molnar, P. *J. geophys. Res.* **86**, 6115–6132 (1981).
- Laubscher, H. *Bull. geol. Soc. Am.* **100**, 1313–1328 (1988).
- Tao, W. C. & O'Connell, R. J. *J. geophys. Res.* **97**, 8877–8904 (1992).
- Davies, H. L. et al. *Contrib. Miner. Petrol.* **103**, 457–469 (1989).
- Saunders, A. D. et al. *Proc. ODP Sci. Res.* **121**, 559–591 (1991).
- Stacey, J. S. & Kramers, J. D. *Earth planet. Sci. Lett.* **26**, 207–221 (1975).
- Ben Othman, D., Polvé, M. & Allègre, C. J. *Nature* **307**, 510–515 (1984).

ACKNOWLEDGEMENTS. We thank J. J. Mahoney and W. M. White for reviews, and A. W. Hofmann, M. McCulloch, W. F. McDonough, K. Mezger, A. E. Ringwood and S.-S. Sun for critical discussions. S.J.G.G. acknowledges support from the Alexander von Humboldt-Stiftung. D. W. is a FNRS Research Associate.

Erythroid transcription factor NF-E2 is a haematopoietic-specific basic-leucine zipper protein

Nancy C. Andrews*, Hediye Erdjument-Bromage†, Mark B. Davidson*, Paul Tempst† & Stuart H. Orkin*^{‡§}

* Division of Hematology/Oncology, Children's Hospital, Dana-Farber Cancer Institute, Department of Pediatrics, Harvard Medical School, 300 Longwood Avenue, Boston, Massachusetts 02115, USA

† Molecular Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA

‡ Howard Hughes Medical Institute, Boston, Massachusetts 02115, USA

Expression of globin genes in developing erythroid cells is controlled by upstream locus control regions. Activity of these regions *in vivo* requires an erythroid-specific nuclear factor (NF-E2) that binds AP-1-like recognition sites. Its tissue-specific component (p45 NF-E2) has been characterized by complementary DNA cloning as a new basic region-leucine zipper protein which dimerizes with a ubiquitous partner to form native NF-E2.

EXPRESSION of globin genes in developing red blood cells is dependent on evolutionarily conserved sequences located upstream of the α - and β -globin complexes, termed locus control regions (LCRs)¹. These regulatory elements coincide with DNase I-hypersensitive domains of erythroid chromatin²⁻⁴ and influence transcription of globin genes more than 50 kilobases (kb) downstream. In transgenic mice, test genes linked to LCRs are consistently expressed in an erythroid-specific manner which is largely independent of the integration site in the host genome³. Naturally occurring deletions of LCR sequences render downstream but intact globin genes inactive in a subset of individuals with α - or β -thalassaemia^{5,6}. Current models postulate interactions between LCRs and proximal regulatory elements of globin genes through the formation of stable complexes assembled by chromatin-bound proteins. These associations presumably change upon switching of globin genes during ontogeny⁷⁻¹⁰. Characterization of the proteins mediating LCR function *in vivo* is fundamental to understanding how LCRs control gene expression.

The active sequence elements of LCRs have been defined through functional assays in transgenic mice and transfected erythroid cells, and by *in vitro* and *in vivo* DNA footprinting methods^{4,11-18}. The principal regulatory element for the human α -globin complex is contained in 300 base pairs (bp) lying 40 kb upstream of the embryonic ζ -globin gene. The β -globin LCR comprises active subdomains of ~300 bp, each encompassing a site of DNase I hypersensitivity (designated HS-1, 2, 3 and 4), located 5-18 kb upstream of the embryonic ϵ -globin gene. In transient transfection assays, the α -globin LCR and β -globin LCR HS-2 sites exhibit classical enhancer activity in erythroid cells¹⁹⁻²², whereas β LCR HS-3 and HS-4 sites do not. Nonetheless, these latter elements contribute substantially to integration site-independent gene expression in transgenic mice. Although the precise structure and organization of individual LCR elements differ, they share a remarkably limited array of motifs recognized by sequence-specific DNA-binding proteins^{11-15,18}. The position-independent property of LCRs correlates with the integrity of CACC/GGTG and GATA motifs. Overall 'enhancer' activity depends largely on TGA(C/G)TCA (AP-1-like) motifs¹³. Ubiquitous proteins bind the CACC/GGTG motifs¹¹ and the abundant erythroid transcription factor GATA-1 (refs 1, 23, 24) recognizes the GATA motifs. Functional analysis of tandem AP-1-like motifs of β LCR HS-2 in both transgenic mice

and transfected cells^{12,13,25,26} indicates that an erythroid-restricted factor, termed NF-E2 (ref. 25), mediates activity of these elements *in vivo*. Available evidence suggests that the ubiquitous Fos and Jun components of AP-1 are not active at these sites in erythroid cells. NF-E2 has been distinguished from AP-1 complex proteins by its faster mobility in gel shift assay, restriction to erythroid and megakaryocytic cells²⁷, absence of Fos/Jun antigenic determinants^{13,26}, and a requirement for a G residue at position -2 (relative to the AP-1 core) for DNA binding²⁸. NF-E2 recognition sites are found in the promoters of two genes encoding enzymes of haem biosynthesis, uroporphobilinogen deaminase²⁸ and ferrochelatase²⁹, as well as in the α - and β -globin LCRs, suggesting a broad role for the protein in erythroid gene expression.

Here we describe the characterization of murine NF-E2 and the molecular cloning of the cDNA encoding its tissue-restricted component. This polypeptide, designated p45 NF-E2, is a new haematopoietic cell-specific member of the basic region-leucine zipper (b-zip) family of transcription factors, and shows striking homology to two invertebrate proteins, encoded by *cnc* and *skn-1*, which are postulated to regulate embryonic development in *Drosophila* and *Caenorhabditis elegans*, respectively^{30,31}. As described in the accompanying paper³², a missense mutation in p45 NF-E2 in microcytic (*mk*) mice provides genetic evidence for the importance of the protein *in vivo*. Expression of recombinant p45 NF-E2 in non-haematopoietic cells is sufficient to generate authentic NF-E2 DNA-binding activity. Thus it should now be possible to introduce both the *cis* and *trans* components for erythroid-specific transcription (specific promoters, LCR core elements, GATA-1 and NF-E2) into heterologous cells so that the mechanistic aspects of LCR function can be addressed.

Binding specificity of NF-E2

NF-E2 was purified from mouse erythroleukaemia (MEL) cells, which contain NF-E2 but little AP-1 activity; activity was followed by gel shift assay.

NF-E2 recognizes a site containing an intact AP-1-binding motif, preceded by a G residue two base pairs upstream²⁸. To characterize this recognition sequence further, we used a series of related oligonucleotides to compete for binding of purified NF-E2 to an NF-E2-site probe (Fig. 1a, b). All oligonucleotides with intact AP-1 motifs {TGA(C/G)TCA} competed for AP-1 binding as expected, independently of their effect on NF-E2 binding (data not shown). In addition, when probes were radio-labelled, only those oligonucleotides that competed for NF-E2

§ To whom correspondence should be addressed at the Children's Hospital.

binding generated complexes with NF-E2 (data not shown). These results suggest that there is an extended consensus sequence for NF-E2 binding: {(T/C)GCTGA(C/G)TCA(T/C)} (Fig. 1b). Oligonucleotides that match this consensus or differ at a single position (outside the AP-1 core) generally bind NF-E2 efficiently; those that differ at two or more positions bind less well, if at all. Although the dyad symmetric AP-1 motif is required, the NF-E2 recognition site is significantly larger and asymmetric.

Two major polypeptides of $M_r \sim 45,000$ that were evident upon electrophoresis of purified NF-E2 (Fig. 2a) were subjected to tryptic cleavage and microsequencing (Fig. 2b, c). Similar HPLC profiles and amino-acid sequences of the tryptic peptides suggested that the doublet bands are encoded by the same gene. cDNA clones were isolated from a MEL cDNA library using degenerate oligonucleotides based on the tryptic peptide sequences. Full-length cDNA and deduced polypeptide sequences are shown in Fig. 3a. An open reading frame extends from a nucleotide 334 to position 1,453, predicting a polypeptide of 373 amino acids (calculated M_r 41.6K), which we designate p45 NF-E2.

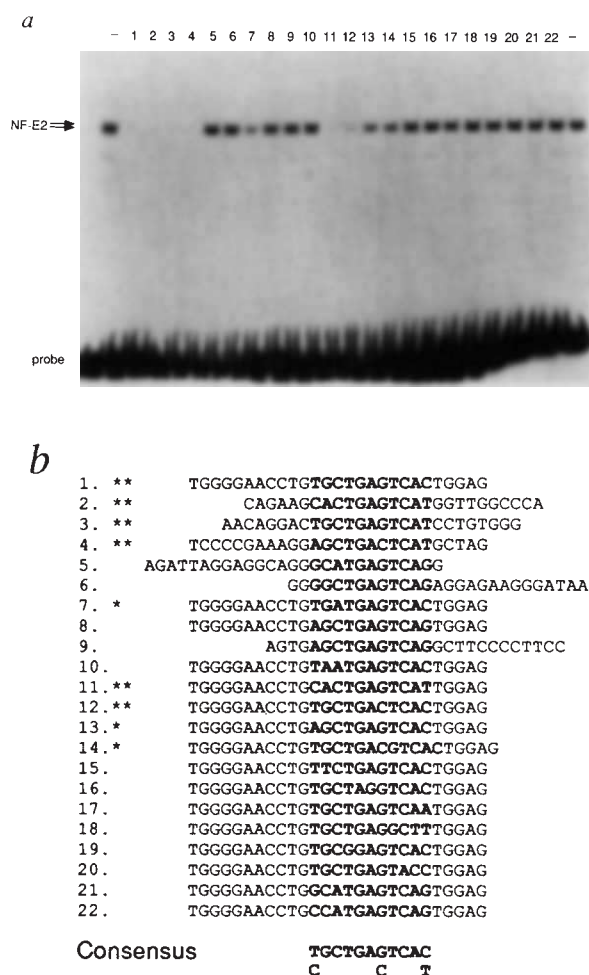


FIG. 1 Consensus recognition site for NF-E2. **a**, Gel-shift competition. Purified murine NF-E2 was incubated with radiolabelled NF-E2-binding-site probe (oligonucleotide 1) in the absence or presence of excess unlabelled competitors, and electrophoresed through a non-denaturing polyacrylamide gel. Arrows indicate the NF-E2-DNA complex. Oligonucleotide competitors are designated 1–22 at the top of the gel; sequences of these competitors are shown in **b**. **b**, Oligonucleotide competitors. Competitive binding seen in **a** is indicated by ** (strong competition) or * (weak competition). A consensus compiled from inspection of sequences that bind efficiently is shown at the bottom.

METHODS. Purified NF-E2 was prepared as described in the legend to Fig. 2. Gel shifts and competitions were performed as described¹⁴.

p45 NF-E2 is a b-zip transcription factor

A striking feature of the predicted p45 NF-E2 polypeptide sequence is a basic potential DNA-binding domain which is followed by heptad repeats of hydrophobic residues, a presumptive leucine zipper³³ (Fig. 3b). The basic region is related to that found in mammalian b-zip proteins of the Jun/Fos and CREB/ATF families^{34–39}, and amino-acid residues that are conserved among these proteins are also conserved in NF-E2 (Fig. 3b; boxed sequences). The basic region of p45 NF-E2 is most similar to that of two presumed invertebrate transcription factors: Cnc encoded by *cnc* of *Drosophila* and Skn-1 encoded by *skn-1* of *C. elegans*^{30,31}. Cnc is developmentally regulated and thought to control head segmentation. Skn-1 is a maternal gene product that acts at the four-cell stage to specify the fate of a blastomere responsible for formation of the pharynx and intestine. The homology to Cnc is striking (Fig. 3c). Unlike the other proteins, Cnc also shares extensive homology with p45 NF-E2 outside the b-zip region, particularly near the carboxy terminus (Fig. 3c).

To establish that p45 NF-E2 cDNA encodes a component of authentic NF-E2, we raised antiserum against the recombinant protein. In western blots of MEL cell extracts, the antiserum specifically recognizes a doublet of ~45K (see below, and Fig. 6a). When added to gel shift assays, anti-p45 NF-E2 antiserum, but not preimmune serum, blocks formation of the complex between native NF-E2 in MEL nuclear extract and a test DNA probe (Fig. 4, left). Antiserum prepared against murine p45 also blocks binding of human NF-E2 to DNA (Fig. 4, right). Formation of AP-1 complexes is not affected by addition of antiserum. These findings demonstrate that the cloned cDNA encodes a component of NF-E2 required for DNA binding, and they also suggest that there is a close structural similarity between mouse and human p45 NF-E2. This latter prediction has been verified by the cloning of cDNA for the human homologue (P. A. Ney *et al.*, manuscript submitted).

Haematopoietic-specific expression of p45 NF-E2

Mammalian b-zip proteins of the Jun/Fos/ATF/CREB family are generally widely expressed. In contrast, NF-E2 binding activity has been reported in only erythroid and megakaryocytic cells²⁷. To investigate whether p45 NF-E2 transcription is tissue-restricted, we examined messenger RNA in various cell lines and mouse tissues. A 1.8-kb RNA transcript is detected in MEL cells (committed erythroid cells), interleukin-3 (IL-3)-dependent haematopoietic progenitor lines (Ba/F3, FDCP-1, 32D), and in fetal liver (Fig. 5). Undifferentiated mouse embryonic stem (ES) cells lack p45 NF-E2 RNA. ES cells differentiated *in vitro* into mature haematopoietic cells, including primitive erythroblasts, express p45 NF-E2 RNA (Fig. 5), whereas those allowed to differentiate along non-haematopoietic pathways do not (data not shown). Of the cell lines tested, all erythroid, haematopoietic progenitor, megakaryocytic and mast cell lines express p45 NF-E2, whereas macrophage, lymphoid and fibroblast lines do not (Fig. 5 legend). Moreover, of the normal tissues examined, only fetal liver, adult spleen and bone marrow, all primary sites of haematopoiesis in the mouse, contain detectable p45 NF-E2. Thus p45 NF-E2 RNA expression mirrors the distribution of NF-E2 DNA-binding activity. This expression pattern is remarkably similar to that for GATA-1 (ref. 40).

In the accompanying report³², we demonstrate that p45 NF-E2 mRNA is also produced in the proximal small intestine of severely anaemic, β -thalassaemic mice (although it is not detectable in intestine RNA of normal mice), and we propose that the protein may play a part in controlling enteric iron absorption. With this single exception, p45 NF-E2 mRNA expression is restricted to haematopoietic tissues and cell lines.

Expression of p45 generates NF-E2 activity

To determine whether p45 alone is sufficient to generate NF-E2 DNA-binding activity in a non-haematopoietic cell, we

introduced transient expression constructs containing either wild-type p45 NF-E2 cDNA, a modified version mutated in the basic region, or antisense-oriented cDNA, into monkey kidney (COS) cells, and assayed nuclear extracts for p45 NF-E2 protein and NF-E2 activity by western blot and gel shift analysis, respectively. As shown in Fig. 6a, p45 NF-E2 protein was detected in cells transfected with wild-type and modified cDNAs, but absent from cells expressing antisense cDNA. NF-E2 DNA-binding activity was found only in transfectants expressing wild-type cDNA (Fig. 6b). This activity is indistinguishable from MEL cell NF-E2 activity in the gel shift mobility of the protein-DNA complex, the oligonucleotide competition profile, and inhibition by anti-p45 NF-E2 antiserum (Fig. 6b, and data not shown). Failure of mutant p45 NF-E2 cDNA to direct NF-E2 DNA-binding activity in COS cells verifies that the basic region is responsible for DNA binding. These experiments establish p45 NF-E2 as the sole cell-specific component required for the formation of NF-E2.

As previously described b-zip proteins bind DNA as dimers, we also asked whether p45 NF-E2 recognizes DNA as a homodimer or as an obligate heterodimer with another component. We therefore coexpressed wild-type and N-terminal-

truncated p45 NF-E2 cDNAs in COS cells and looked for protein-DNA complexes of intermediate mobility. It can be seen from Fig. 6c that no complex indicative of dimer formation between the wild-type and truncated p45 was evident. Thus, p45 NF-E2 does not form homodimers that bind efficiently to the NF-E2 recognition site. From this, and additional findings discussed later, we conclude that native NF-E2 is an obligate heterodimer between a haematopoietic-specific (p45) and a ubiquitous subunit.

Discussion

Globin LCRs comprise the minimal *cis*-regulatory elements sufficient for erythroid-specific gene expression. Information about them helps in the analysis of the DNA-binding sites and proteins required for cell-specific transcription and erythroid differentiation. The activity of LCR elements is believed to reflect the interaction of DNA-binding activities of apparently ubiquitous proteins that recognize the CACC/GGTG motif, and erythroid-restricted proteins, GATA-1 and NF-E2, which bind GATA and AP-1 like motifs, respectively¹. Although the constellation of bound proteins appears remarkably simple, the arrangement of binding sites is a major determinant of LCR

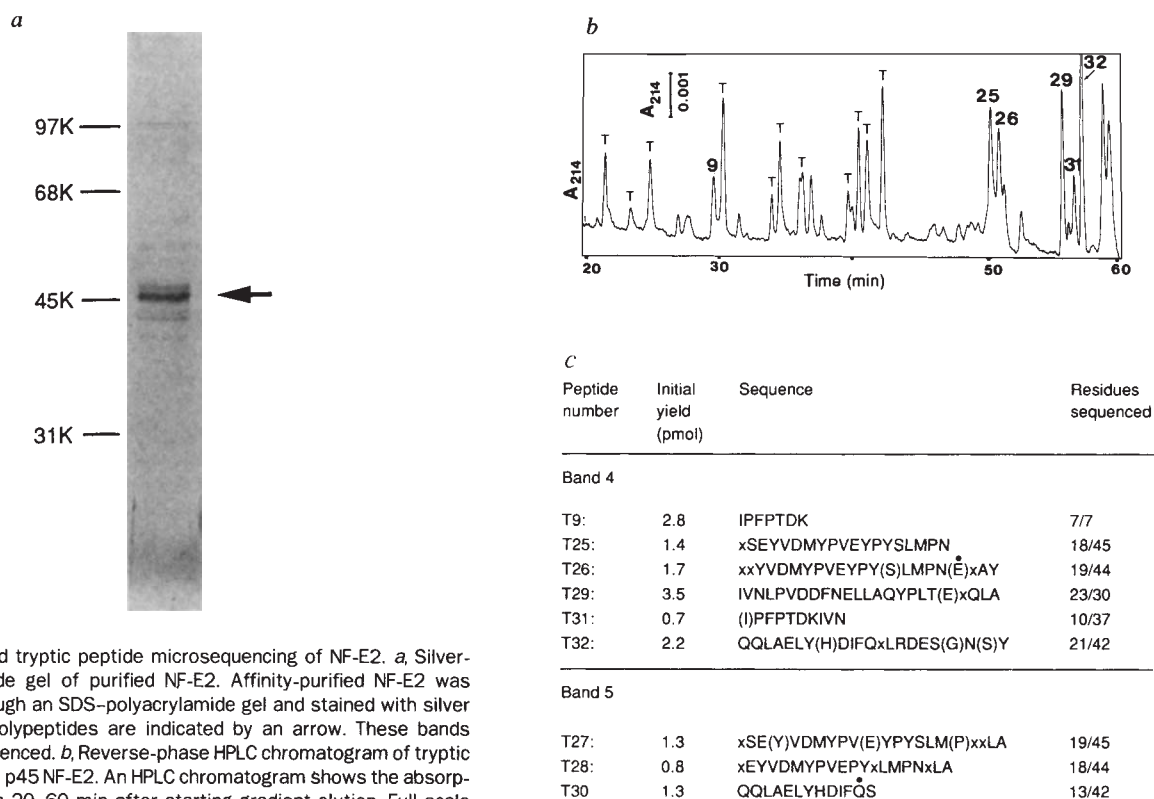


FIG. 2 Purification and tryptic peptide microsequencing of NF-E2. *a*, Silver-stained polyacrylamide gel of purified NF-E2. Affinity-purified NF-E2 was electrophoresed through an SDS-polyacrylamide gel and stained with silver nitrate. Two major polypeptides are indicated by an arrow. These bands were individually sequenced. *b*, Reverse-phase HPLC chromatogram of tryptic peptides derived from p45 NF-E2. An HPLC chromatogram shows the absorption profile at 214 nm 20–60 min after starting gradient elution. Full scale corresponds to 0.005 absorbance units. Peaks resulting from trypsin autolysis are labelled T; numbered peaks correspond to sequenced peptides (see *c*). Comparison of this chromatogram with peptide standard mixtures indicates that peptide peak sizes correspond to ≤ 5 pmol polypeptide. *c*, Amino-acid sequence of NF-E2-derived tryptic peptides. Amino-acid sequences of NF-E2 tryptic peptide peaks are shown in single-letter code. Peptides listed under band 4 correspond to labelled peaks in *b*. Band 5 was derived from the faster band of the 45K NF-E2 doublet. Residues between parentheses are tentative; 'x' indicates that no identification could be made. Calculated initial yields are listed, so are the number of identified amino acids out of the putative total number of residues in that particular peptide. Peptide lengths are predicted from the translated cDNA sequence and the known trypsin specificity, assuming that cleavage of K/R-P and K/R-D/E bonds is very rare⁴⁶. A black dot above an amino acid indicates a discrepancy with cDNA sequencing.

METHODS. NF-E2 DNA-binding activity was purified from 700 g mouse erythroleukaemia cells (MEL clone 745, GM86). Nuclear extract was prepared as

in ref. 47 and fractionated over heparin-Sepharose in Dignam's buffer D with a step gradient of increasing KCl. NF-E2 activity eluted between 0.3–0.4 M KCl. Peak fractions were dialysed to lower salt concentration and passed twice over an NF-E2 site-specific DNA column as in ref. 48. Active material from the final step was precipitated with 0.015% sodium deoxycholate/6% trichloroacetic acid, electrophoresed through a 10% SDS-polyacrylamide gel and stained with silver nitrate⁴⁹. *b*, Active fractions from the final step in NF-E2 purification were precipitated and fractionated by SDS-PAGE, electroblotted onto a nitrocellulose membrane, and visualized by Ponceau S staining. Resulting bands, estimated to contain no more than 0.5 μ g protein, were excised, digested *in-situ* with trypsin (1 μ g sequencing grade trypsin; Boehringer-Mannheim), and the resulting fragments separated by narrow-bore reverse-phase HPLC^{50,51}. *c*, HPLC peak fractions selected for analysis were applied in a model 477A ABI (Foster City, CA) automated sequencer. In view of the exceedingly small initial yields expected, care was taken to optimize the instrument before each run^{52,53}.

a GGAGAGAGGGCCACGGGGAGGAGGGTTGGAGGAGCACTGGTAATCCGACCGAGGGCTTCGGAGAGACCGCCGG 80
 TCGCCCCCTTACCGGGGCAACCGGTGCTCCGCGCCAGCCATGGCCGGGGCGGGGCTCGGCCCTCAGTAGTACT 160
 CAGAGGAGCGCTCAGCGCCAGGAGGTGTCAGCTCTGACCCAGCCTCTCAGGAGACATACCCGACGCTCATCTCTCTC 240
 CCGCGCGGGTTGGAAATCGCCACAGGTGCTGAAAGGTTGCAAGAACTCTTGTGCTTGGAGACCTGGGGCCATCAG 320
 CTGGCAGAGTAGGATCCCGCTCTCTCAGCAGAACAGGTTATCAGACCTGCTGTTGGGAGCTTGGAG 400
 M P F C P F Q Q L N P V S E T S F E
 AGATGGAAGTACTGGCAAGAGATCTGTCATTACTGAGCTGCGGGTCTAAATGTTCCAGTGAACATCTTTTGG 480
 E M E L T W Q E I M S I T E L Q G L N V P S E T S F E
 CCTCAAGCAGCCACCCATACCTGGGCACTGCCACCTCCACATATTGCCCTGTTCAATTCATCAGATGAGGCTT 560
 P Q A P T P Y P G P L P P P T Y C P C S I H P D A G F
 CTCCTTCCCCCATCTTATGAGTCCAGCATCTACTCCCATGTCGCAAGTACCATCTCTATGTAATGTAG 640
 S L P P P S Y E L P A S T P H V P E L P Y S Y G N V
 CTATACAGTGTCAAAGCCACTTACCTTTAGGCTGCTCAATGAGCCCTCCAGACCACTTAGCTCTCTGAGCATT 720
 A I P V S K P L T L S G L L P H V P E L P D L L L D I
 GGGCTCCAGTGGGGCAACCAAGCCCAAGAACAGCAATCTGACTCAGGATTATCCCTCACTACAGTATGTCAGA 800
 G L P V G Q P K P Q E D P E S D S G L S L N Y S D A E
 ATCTCTGAGCTAGAGGTTAGGAGGCTGCCAGGCGGAGGAGGAGTACGTGGACATGTACCCAGTGGAGTATCTT 880
 S L E L E G M E A G R R R S E Y V D M Y P V E Y T T
 CACTTATGCCAATTTTGGCCCTCCCACTATCTTCCACCCACTGAGACACCTTGGCTTAGAGTCACTCTCC 960
 S L M P N S L A H P N Y T L P P T E T P L A L E S S
 GGTCCAGTTCGGCTAAGCTGCTGCTGGGAGGAGGAGTGGGAGCGGAGCCCTGGCCATGAAGATTCC 1040
 G P V R A K P A V R G E A G S R D E R R A L A M K I P
 TTTCCCTACGCAACATAGTTAATGTCGGTGTAGTACTTAAAGTGTGGCAGTATCCGCTAACGGAGAGCC 1120
 F P T D K I E L P V D D F N E L L A Q Y P L T E S
 AGCTGGCTCTAGTTCGGACATCCGTCGACGGGCAAGAACAGGTGGCAGCCAAAAGTGTGCAAGAGAAAGTGGAA 1200
 Q L A L V R D I R R R G K N K V A A Q N C R K R K L E
 ACCATTGTGCGAGTGGAGCGAGAGTGGAGCGGCTGAGCAGTGAAGAGGAGCGGCTTCTCAGAGCCGAGGAGGCTGA 1280
 T I V Q L E R E L E R L S S E R E R L L R A R G E A D
 CCGCACTTGGAGTGGCCCAACAGCTGGCAGAGCTGTACCATGATATTTCCAGATCTTCGGATGAATCTGGCA 1360
 R T L E V M R Q Q L A E L Y D I F Q H L R D E S G
 ACAGTTACTCAGCAGGAATATGTACTGCAACAGGCTGTGATGGTGCACTTTTGGTACCCGCTGGACCAAAAATG 1440
 N S Y S P E E Y V L Q Q A A D G A I F L V P R G T K M
 GAGGCTACAGATTGAGTGGTCTAGAGAGTGGACACTGGTGTAGAGACTCCCTCATCTCCCTTGTATGAAGCA 1515
 E A T D *

b

p45 NF-E2	R	D	I	R	R	G	K	N	K	V	A	A	Q	N	C	R	K	R	K	L	E	T	I	V	Q	L	E	R	E	L	E	R	L	S	S	E	R	E	R	L	L	R	A	R	G	E	A	D	R	T	L	E	V	M					
CNC	R	D	I	R	R	G	K	N	K	V	A	A	Q	N	C	R	K	R	K	L	D	Q	I	L	T	L	E	D	E	V	N	A	V	V	K	R	K	T	Q	L	N	Q	D	R	H	L	E	S	E	R	K	R	I						
SKN-1	R	K	I	R	R	G	K	N	K	V	A	A	R	T	C	R	Q	R	R	T	D	R	H	D	K	M	S	H	Y	I	*																												
c-JUN	K	A	E	R	K	R	I	R	N	R	I	A	A	S	K	C	R	K	R	K	L	E	R	I	A	R	L	E	E	K	V	K	T	L	K	A	Q	N	S	E	L	A	S	T	A	N	H	L	R	E	Q	V	A	Q	L				
Jun B	K	V	E	R	K	R	I	R	N	R	I	A	A	T	K	C	R	K	R	K	L	E	R	I	A	R	L	E	D	K	V	K	T	L	K	A	E	N	A	G	L	S	S	A	G	I	L	R	F	Q	V	A	Q	L					
GCN4	P	A	A	L	K	R	A	R	N	T	E	A	A	R	R	S	R	A	R	K	I	Q	R	H	K	Q	L	E	D	K	V	E	L	L	S	K	N	Y	H	L	S	S	E	V	A	R	L	K	K	I	V	G	E	R					
FOS	K	R	R	I	R	R	E	R	N	K	H	A	A	K	C	R	N	R	R	R	E	L	T	D	T	L	Q	A	E	T	D	Q	L	E	D	E	K	S	A	L	Q	T	E	I	A	N	L	K	E	K	F	K	L						
CREB	R	R	R	V	R	H	E	R	N	K	L	A	A	K	C	R	N	R	R	K	E	I	T	D	F	L	Q	A	E	T	D	K	L	E	D	E	K	S	G	L	Q	R	E	I	E	L	K	Q	K	E	R	L							
CREB	K	R	E	V	R	L	M	K	N	R	E	A	R	E	C	R	R	K	K	E	Y	V	K	C	L	E	N	R	V	A	V	L	E	N	Q	N	K	T	L	I	E	E	I	K	A	L	K	D	L	Y	C	H	K						
ATF1	K	R	E	I	R	L	M	K	N	R	E	A	R	E	C	R	R	K	K	E	Y	V	K	C	L	E	N	R	V	A	V	L	E	N	Q	N	K	T	L	I	E	E	L	K	T	L	K	D	L	Y	S	N	K						

----- Basic region ----- Leucine repeats -----

FIG. 3 Sequence of p45 NF-E2 and homology to b-zip proteins. **a**, Nucleotide sequence of cDNA encoding p45 NF-E2, with the deduced amino-acid sequence underneath. Sequences corresponding to tryptic peptides shown in Fig. 2 are indicated in bold. The Genbank accession number for this sequence is L09600. **b**, Sequence alignment of the basic region-leucine zipper domains of p45 NF-E2 and related proteins. The b-zip domain of p45 NF-E2 is shown in bold. Proteins sharing homology in this region are aligned beneath. Boxes indicate the most highly conserved residues of the basic regions, and the heptad repeats of the leucine zippers. Sequences of the b-zip proteins are derived from refs 30, 31, 34-39, 54. **c**, Homology of p45 NF-E2 to *Drosophila* Cnc. Amino-acid sequences from p45 NF-E2 (residues 206-373) and Cnc³⁰ (residues 287-454) are aligned for comparison. Amino-acid identity is indicated by connecting lines; conservation is indicated by dots. The basic region-leucine zipper (b-zip) domains are indicated by a plus sign above basic amino acids, and an asterisk above hydrophobic heptad-repeat amino acids.

METHODS. Degenerate oligonucleotides representing all possible codons of contiguous portions of tryptic peptides T27 and T30 were used to screen a MEL cell plasmid cDNA library prepared in pCDM8 (gift from L. Zon) using standard procedure⁴⁹. Six clones containing the same 1.5-kb insert were isolated from $\sim 5 \times 10^5$ colonies screened. The insert was subcloned for DNA sequencing⁴⁹.

c

NF-E2	206	SSSGPVRAKP	AVRGEAGSRD	ERRALAMKIP	FPTDKIVNLP	VDDFNELLAQ
Cnc	287	SSSVGGNNSN	L-EEHLTRD	EKRARSINLP	ISVPDILNLF	MDPNERLSK

NF-E2	YPLTESQAL	VRDVRGKN	KVAAQNCRRK	KLETIVQLER	ELERLSSERE
Cnc	YDLSENQSL	IRDIRRGKN	KVAAQNCRRK	KLDQITL	EDVNAVKKRT

NF-E2	RLLRARGEAD	RTLEVNRQQL	AELYHDFQH	LRDESGNSYS	PEEYVLQAA
Cnc	QLNQDRDHL	SEKRSINFK	AMLRHVFOY	LRDEGNPCW	PADYSLQAA

NF-E2	DGAIFLVPRG	TKMEATD*
Cnc	DGSVYLPRP	KSEGNNTA

activity¹². Moreover, other nuclear proteins that do not directly contact DNA may interact with LCR-bound factors and contribute to LCR function. Nonetheless, as the major *in vivo* activity of the LCR depends on NF-E2 (refs 11-13), defining its structure and identifying potential interactions between NF-E2 and other transcription factors, such as GATA-1, are at the heart of determining how LCRs function as dominant regulatory elements.

The purification of NF-E2 presented formidable difficulties owing to its low abundance in erythroid cells and the collection of widely distributed Jun/Fos/ATF/CREB family members likely to copurify with it because of their similar DNA-binding properties. To overcome these obstacles, we chose an erythroid

cell-line with a low level of AP-1-like binding activity, and characterized the NF-E2 recognition site extensively to ensure that highly enriched fractions contained authentic NF-E2.

Several lines of evidence establish that the protein encoded by p45 NF-E2 is a major component of NF-E2 DNA-binding activity. First, antiserum against recombinant protein blocks formation of the NF-E2-DNA complex, but does not disrupt complexes containing AP-1 proteins (Fig. 4). Second, forced expression of p45 NF-E2 in non-haematopoietic cells is sufficient to generate NF-E2 DNA-binding activity (Fig. 6). Third, the cellular distribution of p45 NF-E2 expression precisely mirrors that of NF-E2 binding activity (Fig. 5). Thus, the protein encoded by p45 NF-E2 is derived from native NF-E2, rather than from an unrelated b-zip protein with fortuitously similar properties. Furthermore, the results reported in the accompanying letter³², which demonstrate a missense mutation in the p45 NF-E2 gene in the microcytic *mk* mouse, provide genetic evidence for its important role in the regulation of both globin gene expression and iron metabolism.

p45 NF-E2 is distinctive for its restricted cellular distribution. Just as the NF-E2 DNA-binding complex is restricted to a subset of haematopoietic cell types, the NF-E2 recognition sequence appears to be preferentially associated with regulatory elements for erythroid-cell gene expression. Although AP-1 and ATF/CREB recognition sites commonly appear in the promoters and enhancers of diverse genes, a computer search suggests that most sites matching the 11-bp NF-E2 consensus are

found in regulatory elements of erythroid genes (M. Fleming and N.C.A., unpublished results). Of AP-1 like sites identified in erythroid regulatory elements, purified NF-E2 binds efficiently to sites in the α - and β -globin LCRs, the chicken 3'- β -globin enhancer, and the human porphobilinogen deaminase promoter²⁵. Consensus sites are also found in the promoters for human ferrochelatase²⁹ and chicken H ferritin⁴¹. Non-consensus AP-1-like sites found in the promoters for human aminolevulinic acid synthase (GGCTGAGTCAG)⁴² and mouse band 3 (GCATGAGTCAG)⁴³ do not efficiently bind NF-E2 *in vitro* (Fig. 1). The selective use of the NF-E2 recognition site contrasts with the more pervasive involvement of the GATA motif, which is presented in regulatory elements of virtually all characterized erythroid-expressed genes, as well as in all LCR cores.

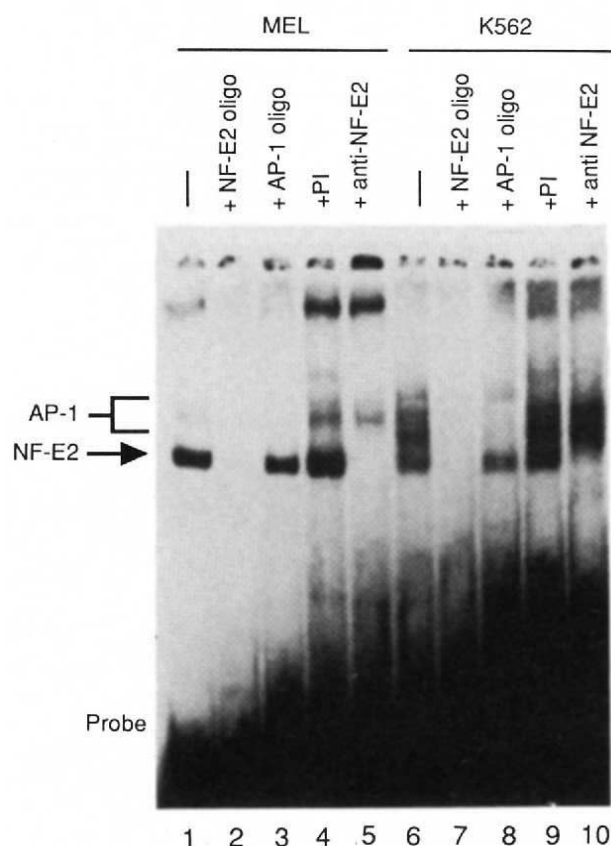


FIG. 4 Antiserum prepared against recombinant p45 NF-E2 specifically blocks NF-E2 binding to DNA. Nuclear extracts from MEL (lanes 1–5) or K562 (lanes 6–10) cell lines were incubated with radiolabelled oligonucleotide-1 probe. Migration of free probe, NF-E2 DNA-binding activity, and AP-1 DNA-binding activity is indicated. To confirm the binding specificity of NF-E2 to DNA, excess unlabelled oligonucleotide competitors were added in lanes 2, 3, 7 and 8 as indicated. Lanes 4, 9 and 5, 10 show reactions to which either preimmune (PI) serum or anti-recombinant p45 NF-E2 serum (anti-NF-E2) was added, respectively.

METHODS. Gel shifts were carried out as described^{14,56}. Either 10 ng unlabelled oligonucleotide competitor or 1 μ l of serum was added in lanes 2–5 and 7–10 before addition of radiolabelled probe, as indicated. Competitor oligonucleotides were designated NF-E2 oligo (TGGGGAACCTG-TGCTGAGTCACTGGAG) and AP-1 oligo (TGGGGAACCTGTTCTGAGTCACTGGAG), corresponding to oligonucleotide competitors 1 and 15 in Fig. 1b. Preimmune and anti-p45 NF-E2 antisera were prepared from the same female New Zealand white rabbit before and after immunization with a glutathione S-transferase (GST) fusion protein⁵⁵ which contained amino acids 1–348 of p45 NF-E2. An unrelated antibody with anti-GST reactivity did not affect the NF-E2 gel shift (data not shown).

Our experiments indicate that native NF-E2 is an obligate heterodimer of p45 NF-E2 and a ubiquitous component. First, by comparing the mobility of the NF-E2 DNA complex with complexes containing AP-1 and GATA-1 (ref. 44), we estimate the NF-E2 M_r to be 60K (data not shown). This estimate is appreciably larger than the M_r of p45 alone, yet considerably smaller than that expected for a p45 homodimer. Second, in cotransfection experiments we failed to detect evidence of a homodimer capable of binding the NF-E2 recognition sequence (Fig. 6c). Following these observations, we re-examined preparations of purified NF-E2 and identified an 18K species present in apparent stoichiometry with the p45 NF-E2 doublet (N.C.A. and S.H.O., unpublished result). Accordingly, an 18K polypeptide co-immunoprecipitates with p45 from metabolically labelled erythroid cells (P. A. Ney *et al.*, manuscript submitted). We infer that the presumptive 18K partner must be widely expressed, as non-haematopoietic cells transfected with expressible p45 cDNA generate NF-E2 DNA-binding activity. The low

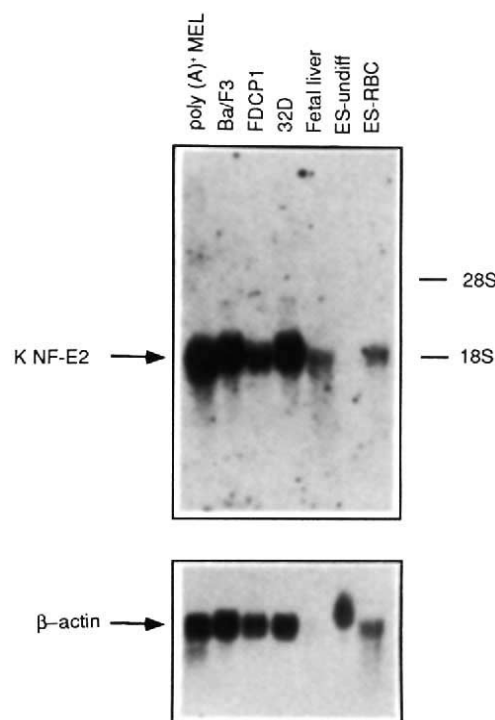
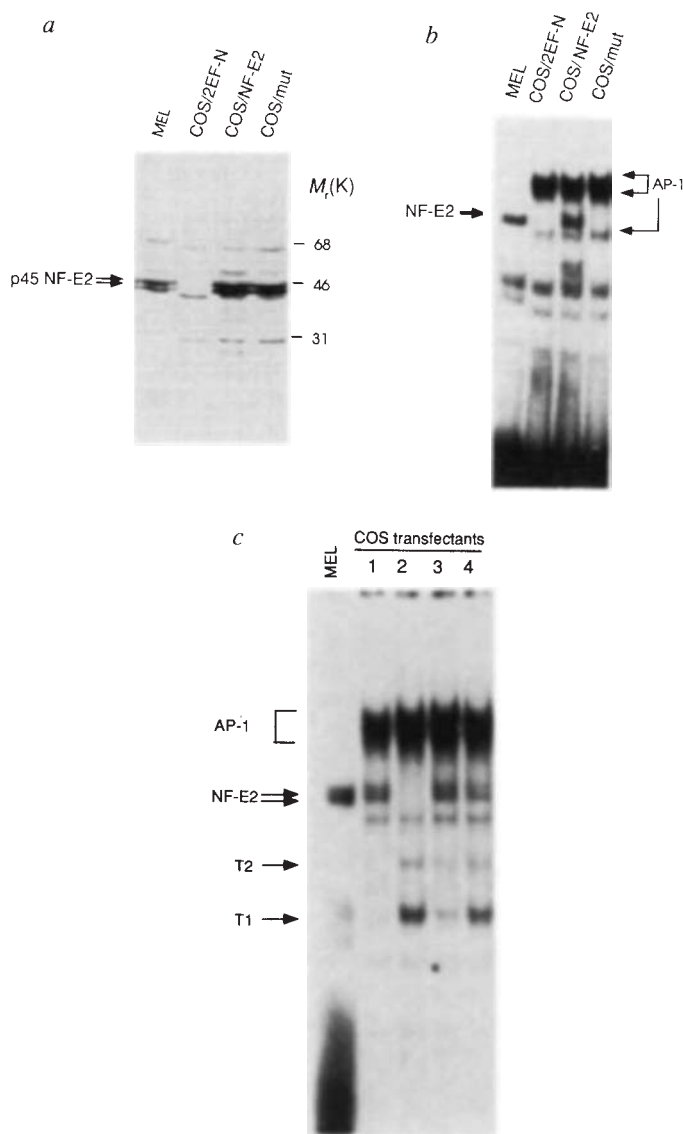


FIG. 5 Expression of p45 NF-E2 RNA in progenitor cell lines, murine fetal liver and differentiated ES cells. Top panel shows hybridization with a probe corresponding to the entire NF-E2 cDNA; lower panel shows rehybridization with a mouse β -actin cDNA probe, as a control for RNA integrity. Ba/F3, FDCP1 and 32D are IL-3-dependent multipotential cell lines. 'ES-undiff' indicates RNA of CCE ES cells; ES-RBC is an RNA sample of ES cells differentiated *in vitro* into erythroid cells. In addition, the following cell lines were positive for p45 NF-E2 RNA expression by northern blot analysis: B6suta (multipotential progenitor); LIN 011, LIN 175, LIN 192 and LIN 261 (megakaryocytic lines); MC8 (mast cell line); the following cell lines were negative: J774 and M1/MT (macrophage lines); NIH 3T3 and LTK (fibroblasts); 38B9, SCID24, SCID44, 3001pp, 18-8, HAFT, 22D7, 28C3, BASC, 70DZ (pre-B cell lines); Whi231, A2002J, Whi239, BCL1, M12 (B-cell lines); MOPC11 and MOPC21 (plasmacytoma lines); T3 and T7 (pre-T cell lines); and EL4 (T-cell line). Of primary mouse tissues, spleen and bone marrow were positive; heart, brain, lung, liver, skeletal muscle, kidney, testis and large intestine were negative. As reported³², p45 NF-E2 RNA was detected in the proximal small intestine of β -thalassaemic mice.

METHODS. Northern blots were prepared by standard procedures and used to assay 2.5 μ g poly(A)⁺ MEL RNA and 20 μ g total RNA from progenitor cell lines, fetal liver and ES cells. Other RNA samples analysed were 2–3 μ g poly(A)⁺ RNA from MEL, MC8, NIH 3T3, heart, brain, spleen, lung, liver, skeletal muscle, kidney or testis, or 20–30 μ g total RNA from remaining cell lines and tissues.

FIG. 6 Reconstitution of NF-E2 DNA-binding activity in transfected COS cells. *a*, Western blot showing 45K bands detected with anti-p45 NF-E2 antiserum in nuclear extracts from MEL cells, and from monkey kidney COS cells transfected with pXM vector expression constructs carrying wild-type p45 NF-E2 cDNA in reverse (2EF-N) or sense (NF-E2) orientation, or a mutant version with four amino-acid substitutions in the basic domain (mut). Equal volumes of nuclear extract were loaded for MEL and COS transfectant samples. *b*, A gel shift experiment is shown in which the extracts in *a* were incubated with a radiolabelled NF-E2 binding-site probe. In this case, the volume of MEL extract used was 1/10 that of the COS transfectant extracts. Free probe was electrophoresed off the bottom of the gel. The position of the endogenous MEL NF-E2 gel shift, and NF-E2 gel shift complex generated by expression of p45 NF-E2 cDNA in COS cells is indicated. AP-1 binding activity endogenous to COS cells is also indicated. *c*, COS cells were transfected with 5 μ g wild-type pXM p45 NF-E2 cDNA (lane 1), 5 μ g pXM vector carrying a truncated form of cDNA (lane 2), 5 μ g wild-type plus 0.5 μ g truncated cDNA (lane 3), or 5 μ g wild-type plus 1 μ g truncated cDNA (lane 4). Arrows indicate the endogenous COS AP-1 gel-shift (AP-1) and the endogenous MEL and full-length COS-transfected NF-E2 gel-shift complexes (NF-E2). T1 and T2 correspond to major and minor complexes generated by transfection of the truncated cDNA expression construct. Differing ratios of wild-type and truncated cDNA plasmids were tested, as the gel-shift signal from the expressed truncated protein was consistently more intense than that of the wild-type protein when equal amounts were cotransfected, suggesting that the truncated protein either had an increased affinity for the DNA-binding site, or that it enhanced dimerization with a ubiquitous partner protein. No new gel-shift complexes are apparent upon cotransfection of the wild-type and truncated constructs, indicating that the expressed proteins cannot dimerize with each other to generate NF-E2 DNA-binding activity.

METHODS. The constructs used in COS transfection experiments contained wild-type or modified NF-E2 cDNA in the mammalian expression vector pXM²⁴; the mutated (mut) construct had 4 amino-acid changes in the basic region (Lys 275 to Glu; Asn 276 to Tyr; Lys 277 to Glu; Ala 279 to Thr), created by site-directed mutagenesis⁴⁹. The truncated form of NF-E2 cDNA used in *c* encoded only the C-terminal 141 amino acids. Plasmids were introduced into COS cells by the DEAE-dextran method⁴⁹. Nuclear extracts were prepared as described⁵⁶. The western blot in *a* was developed with reagents purchased from Promega according to their instructions.



level of binding activity in transfected COS cells (relative to the abundance of expressed p45 polypeptide) further indicates that the partner may be limiting in different cellular environments. We speculate that the asymmetry of the NF-E2 DNA-binding site reflects asymmetry of the NF-E2 heterodimer, as predicted from its subunit composition.

It is notable that GATA-1 and NF-E2, factors studied mainly for their involvement in erythroid gene expression, are both present in haematopoietic progenitors before terminal erythroid commitment and are found in two other haematopoietic cell types, megakaryocytes and mast cells. Their co-expression in three lineages, all descendants from a common, committed progenitor, underscores the close relationship between these cell types^{27,40}. Moreover, co-expression of these factors suggests that they may be coordinately or cross-regulated. As GATA-1 and p45 NF-E2 genes are both expressed in progenitor cells and critical for subsequent erythroid gene expression, dissecting the mechanisms responsible for their regulation is a logical path to defining the earliest events of an erythroid program of cellular development and potential regulatory networks.

GATA-1 and NF-E2 appear to be the major cell-restricted, DNA-binding proteins that interact directly with the DNA of the LCR cores. Hence, the distinctive properties of LCRs in enhancing erythroid cell gene expression are likely to result from interaction of these factors with each other, and with other components of the proposed LCR-promoter complexes. A para-

dox in our understanding of LCR function remains, however. Of the cell types expressing both GATA-1 and NF-E2 (erythroid cells, haematopoietic progenitors, megakaryocytes, and mast cells), only erythroid cells actively transcribe globin genes. Whether critical, cell-specific post-translational modifications and/or essential accessory factors regulate the function of these proteins in different cellular contexts remains unknown.

Characterization of NF-E2 and molecular cloning of the cDNA encoding its haematopoietic-specific component are significant steps towards an analysis of LCR function. We are now in a position to examine how the cell-specific LCR-binding proteins (GATA-1 and NF-E2) assemble on DNA and interact with each other and ubiquitous LCR-bound factors, as well as proximal promoter-bound proteins, to activate gene expression. With gene targeting in mouse embryonic stem cells and the methods available for *in vitro* differentiation of haematopoietic cells, it will be possible to initiate a formal genetic analysis of the function of NF-E2 in erythroid cell development⁴⁵. One challenge will be to define the unique properties of NF-E2 that enable it to serve as a major enhancer protein of LCRs. □

Received 1 December 1992; accepted 23 February 1993.

- Orkin, S. H. *Cell* **63**, 665-672 (1990).
- Tuan, D., Solomon, W., Li, Q. & London, I. M. *Proc. natn. Acad. Sci. U.S.A.* **82**, 6384-6388 (1985).
- Grosfeld, F., van Assendelft, G. B., Greaves, D. R. & Kollias, B. *Cell* **51**, 975-985 (1987).
- Higgs, D. R. *et al. Genes Dev.* **4**, 1588-1601 (1990).

5. Van der Ploeg, L. H. T. *et al.* *Nature* **283**, 637–642 (1980).
6. Hatton, C. S. R. *et al.* *Blood* **76**, 221–227 (1990).
7. Behringer, R. R., Ryan, T. M., Palmiter, R. D., Brinster, R. L. & Townes, T. M. *Genes Dev.* **4**, 380–389 (1990).
8. Enver, T. *et al.* *Nature* **344**, 309–313 (1990).
9. Choi, O.-R. & Engel, J. D. *Cell* **55**, 17–26 (1988).
10. Hanscombe, O. *et al.* *Genes Dev.* **5**, 1387–1394 (1991).
11. Philipsen, S., Talbot, D., Fraser, P. & Grosfeld, F. *EMBO J.* **9**, 2159–2167 (1990).
12. Talbot, D., Philipsen, S., Fraser, P. & Grosfeld, F. *EMBO J.* **9**, 2169–2178 (1990).
13. Talbot, D. & Grosfeld, F. *EMBO J.* **10**, 1391–1398 (1991).
14. Strauss, E. C., Andrews, N. C., Higgs, D. R. & Orkin, S. H. *Molec. cell. Biol.* **12**, 2135–2142 (1992).
15. Strauss, E. C. & Orkin, S. H. *Proc. natn. Acad. Sci. U.S.A.* **89**, 5809–5813 (1992).
16. Ikuta, T. & Kan, Y. W. *Proc. natn. Acad. Sci. U.S.A.* **88**, 10188–10192 (1991).
17. Reddy, P. M. S. & Shen, C.-K. *Proc. natn. Acad. Sci. U.S.A.* **88**, 8676–8680 (1991).
18. Jarman, A. P. *et al.* *Molec. cell. Biol.* **11**, 4679–4689 (1991).
19. Tuan, D. Y., Solomon, W. B., London, I. M. & Lee, D. P. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2554–2558 (1989).
20. Ney, P. A., Sorrentino, B. P., McDonagh, K. T. & Nienhuis, A. W. *Genes Dev.* **4**, 993–1006 (1990).
21. Pondel, M. D., George, M. & Proudfoot, N. J. *Nucleic Acids Res.* **20**, 237–243 (1992).
22. Moon, A. M. & Ley, T. J. *Blood* **77**, 2272–2284 (1991).
23. Evans, T. & Felsenfeld, G. *Cell* **58**, 877–885 (1989).
24. Tsai, S. F. *et al.* *Nature* **339**, 446–451 (1989).
25. Mignotte, V., Wall, L., deBoer, E., Grosfeld, F. & Romeo, P.-H. *Nucleic Acids Res.* **17**, 37–54 (1989).
26. Ney, P. A., Sorrentino, B. P., Lowrey, C. H. & Nienhuis, A. W. *Nucleic Acids Res.* **18**, 6011–6017 (1990).
27. Romeo, P.-H. *et al.* *Nature* **344**, 447–449 (1990).
28. Mignotte, V., Eleouet, F., Raich, N. & Romeo, P.-H. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6548–6551 (1989).
29. Taketani, S., Inazawa, J., Nakahashi, Y., Abe, T. & Tokunaga, R. *Eur. J. Biochem.* **205**, 217–222 (1992).
30. Mohler, J., Vani, K., Leung, S. & Epstein, A. *Mech. Dev.* **34**, 3–10 (1991).
31. Bowerman, B., Eaton, B. A. & Priess, J. R. *Cell* **68**, 1061–1075 (1992).
32. Peters, L. L. *et al.* *Nature* **362**, 768–770 (1993).
33. Landschultz, W. H., Johnson, P. F. & McKnight, S. L. *Science* **240**, 1759–1764 (1988).
34. Bohmann, D. *et al.* *Science* **238**, 1386–1392 (1987).
35. Cohen, D. R. & Curran, T. *Molec. cell. Biol.* **8**, 2063–2069 (1988).
36. Hai, T., Liu, F., Coukos, W. J. & Green, M. R. *Genes Dev.* **3**, 2083–2090 (1989).
37. Hinnebusch, A. G. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6442–6446 (1984).
38. Hoeffler, J. P., Meyer, T. E., Yun, Y., Jameson, J. L. & Habener, J. F. *Science* **242**, 1430–1433 (1988).
39. Van Straaten, F., Muller, R., Curran, T., van Beveren, C. & Verma, I. M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3183–3187 (1983).
40. Martin, D. I. K., Zon, L. I., Mutter, G. & Orkin, S. H. *Nature* **344**, 444–446 (1990).
41. Wilkins-Stevens, P., Dodgson, J. B. & Engel, J. D. *Molec. cell. Biol.* **7**, 1751–1758 (1987).
42. Cox, T. C., Bawden, M. J., Martin, A. & May, B. K. *EMBO J.* **10**, 1891–1902 (1991).
43. Kopito, R. R., Anderson, M. A. & Lodish, H. F. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7149–7153 (1987).
44. Park, C. W. & Walker, M. D. *J. Biol. Chem.* **267**, 15642–15649 (1992).
45. Simon, M. C. *et al.* *Nature Genet.* **1**, 92–98 (1992).
46. Tempst, P. & Van Beeumen, J. *Eur. J. Biochem.* **135**, 321–330 (1983).
47. Dignam, J. D., Martin, P. L., Shastry, B. S. & Roeder, R. G. *Meth. Enzym.* **101**, 587–597 (1983).
48. Kadonaga, J. T. *Meth. Enzym.* **208**, 10–23 (1991).
49. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* 2nd edn (Cold Spring Harbor Laboratory Press, New York, 1989).
50. Tempst, P., Link, A. J., Riviere, L. R., Fleming, M. & Ellicone, C. *Electrophoresis* **11**, 537–553 (1990).
51. Ghosh, S. *et al.* *Cell* **62**, 1019–1029 (1990).
52. Tempst, P. & Riviere, L. *Analyt. Biochem.* **183**, 290–300 (1989).
53. Erdjument-Bromage, H., Geromanos, S., Chodera, A. & Tempst, P. in *Techniques in Protein Chemistry IV* (ed. Hogue-Angeletti, R.) (Academic, San Diego, in the press).
54. Ryder, K., Lau, L. F. & Nathans, D. *Proc. natn. Acad. Sci. U.S.A.* **85**, 1487–1491 (1988).
55. Blackwood, E. & Eisenman, R. N. *Science* **251**, 1211–1217 (1991).
56. Andrews, N. C. & Faller, D. V. *Nucleic Acids Res.* **19**, 2499 (1991).

ACKNOWLEDGEMENTS. Large scale cultures of MEL cells and primary nuclear extracts were produced by Paragon Biotech, Baltimore, under the supervision of M. Chacon. We thank S. Goff for synthesis of oligonucleotides, L. Zon for the pCDM8 MEL cDNA library and for megakaryocyte cell line RNAs, P. Richardson for MC-8 RNA, M. C. Simon for ES cell RNAs, B. Matthey-Prevot for progenitor cell line RNAs, X. Yang for lymphoid and macrophage RNAs, E. Whitelaw for fetal liver RNA, and S. Lux for assistance with sequence comparisons and for encouragement. This work was supported by a Merck-AFCR foundation MD-PhD postdoctoral fellowship to N.C.A. and an NIH grant to S.H.O., who is an Investigator of the Howard Hughes Medical Institute.

LETTERS TO NATURE

Recurrent burst activity from the soft γ -ray repeater SGR 1900+14

C. Kouveliotou[†], G. J. Fishman^{*}, C. A. Meegan^{*}, W. S. Paciesas[‡], R. B. Wilson^{*}, J. van Paradijs[§], R. D. Preece^{*}, M. S. Briggs[‡], G. N. Pendleton[‡], M. N. Brock^{*}, T. M. Koshut[‡] & J. M. Horack^{*}

^{*} Space Science Laboratory, ES 66, NASA/Marshall Space Flight Center, Huntsville, Alabama 35812, USA

[‡] Department of Physics, University of Alabama in Huntsville, Alabama 35899, USA

[§] Astronomical Institute 'Anton Pannekoek' and Center for High-Energy Astrophysics, Kruislaan 403, 1098 SJ Amsterdam, The Netherlands

OF the one thousand or so detected sources of cosmic γ -ray bursts (GRBs), only three are known to have exhibited recurrent activity^{1–4}. These events differ from the majority of GRBs in that they are of much shorter duration and have relatively soft spectra⁵. The recurrent sources can thus be considered as a distinct class of objects—the soft gamma repeaters (SGRs). The embryonic distribution of the SGRs suggests that they are of galactic origin⁴, whereas the distribution⁶ of GRBs seems to be inconsistent with any known galactic source population. Here we report the detection by the Burst and Transient Source Experiment (BATSE)⁷ of three short, very soft transient events from a location consistent with that of the 'old' repeater SGR 1900+14 (ref. 8). Our results suggest that the SGR active phase lasts at least 13 years, which, in conjunction with the arguments for a galactic origin, lends support to suggestions^{4,5} that SGRs are neutron stars.

The three SGRs were discovered with the KONUS experiment on the Venera 11–14 spacecraft⁹ and with the first and second Interplanetary Networks (IPN) of satellites^{1–4}. With one exception (the initial outburst of SGR0526–66¹⁰, the famous 5 March 1979 event) all emissions from these sources have unique and

rather uniform characteristics⁵: their durations are extremely short (typically ~ 0.1 s), their light-curves are simple (triangular or 'flat-topped'), and their rise times can be as short as 5 ms (ref. 4). Their spectra are much harder than those of X-ray bursts¹¹ and much softer than those of classical GRBs^{12,13}; they are well fitted with an optically thin thermal bremsstrahlung (OTTB) function with temperatures between 30 and 40 keV (refs 4, 14) and show no evidence for spectral evolution⁴. The total numbers of events so far observed from SGR0526–66, SGR1900+14 and SGR1806–20 are 16, 3 and over 100, respectively, with intervals between events that vary stochastically from hours to years.

The first SGR event detected with BATSE⁷ (Fig. 1a and b) consists of two pulses separated by ~ 0.5 s. The first pulse (Fig. 1a) lasts 40 ms and has a trapezoidal shape. It initially rises

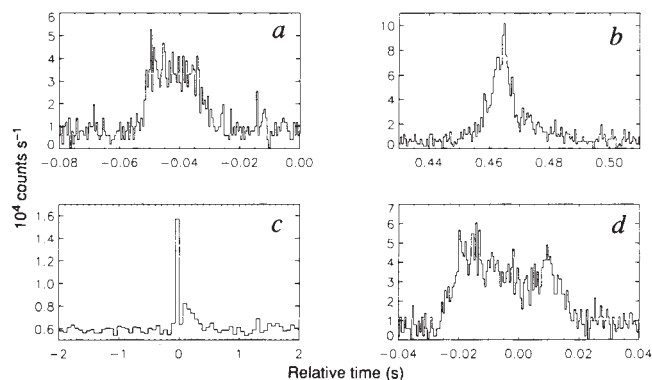


FIG. 1 Light curves of the three SGR bursts detected with the Burst and Transient Source Experiment (BATSE) on the Compton Gamma Ray Observatory. BATSE is unique in combining good sensitivity to weak γ -ray and SGR-like bursts with the capability of determining their locations in a single experiment. *a, b*, First and second pulse, respectively, of the 19 June 1992 trigger. The counts are integrated between 20 and 100 keV. Time resolution is 0.512 ms. *c*, Event of 8 July 1992 recorded with 64-ms time resolution (the only resolution available). The initial spike of <64 ms appears only up to 100 keV. *d*, Event of 19 August 1992 recorded with 0.512-ms resolution between 20 and 100 keV. Notice the similarity in structure between *a* and *d*.

[†] On leave from the University of Athens, Greece.

gradually for ~ 10 ms and then increases sharply in ~ 0.5 ms to a plateau lasting ~ 20 ms; it then decays to near the background level in less than 5 ms with some residual weak emission continuing above background for the following 30 ms. These rise and decay times are the shortest yet resolved for a SGR event (they are not limited by the instrument resolution of $0.5 \mu\text{s}$). They may reflect the true onset and cessation of the physical process involved and provide an upper limit on the size of the emitting region (~ 150 km). As a comparison, the unresolved rise time of the 5 March 1979 event was about a factor of two (or more) faster. The second pulse (Fig. 1b) also lasts 40 ms and is triangular. It rises gradually in 15 ms, and then decays to the background level in 25 ms. Neither of the two pulses shows significant substructure. The second SGR burst (Fig. 1c) lasts ~ 500 ms with most of the emission concentrated in an unresolved initial spike lasting less than 64 ms. The third SGR event (Fig. 1d) lasts ~ 80 ms and has a trapezoidal (flat-topped) shape with a plateau lasting ~ 36 ms, a fast rise of < 5 ms and a more gradual decay of ~ 40 ms.

We have sufficient data for deriving a spectrum only for the second pulse of the first event. Figure 2 shows the best fit of an OTTB function to the time-integrated spectrum of this pulse, with $kT = 39 \pm 3$ keV ($\chi^2 = 10.7$ for 11 degrees of freedom). A power-law fit has a spectral index of -3.0 ± 0.1 but a significantly worse χ^2 (29.8 for 12 degrees of freedom). Spectral fits for two independent 16-ms time bins show no spectral evolution during the pulse.

Some spectral information for the remaining three pulses can be obtained from their hardness ratios R_{21} and R_{32} , which are the ratios of their background-corrected count rates in the 50–100 to 20–50 keV bands and 100–300 to 50–100 keV bands, respectively (see Table 1). The hardness ratios of the four pulses are similar (the first pulse of the 19 June event is slightly harder). The time histories (in 1-ms time bins) of the hardness ratios of three of the pulses from Fig. 1 (a, b and d) do not reveal a significant spectral change during any of these events. The ratios R_{21} are much larger than R_{32} ; this is the reverse of what is seen for classical γ -ray bursts¹⁵ and is another confirmation of the distinctly different nature of these two types of high-energy bursts. This shows that there is no observational ambiguity about BATSE's capability to detect or trigger on SGRs, and firmly puts to rest a theoretical speculation¹⁶ that confusion between SGRs and GRBs would be responsible for the observed⁶ inhomogeneity of GRBs.

We have used the spectral index of -3.0 as an initial estimate of the input power-law spectrum required to compute the locations of the events. All three locations are consistent with a single positional error circle in the sky centred at $\alpha = 288^\circ$, $\delta = 11^\circ$, and 1σ radius $\sim 5^\circ$. Because no other spacecraft detected these emissions (K. Hurley, personal communication), a much more accurate location from triangulation cannot be obtained. This error circle is consistent with the location of the 2σ error box of SGR1900+14 (see Fig. 3), which was obtained by triangulation of the KONUS experiments (E. Mazets, personal communication). In view of the temporal and spectral similarities between the BATSE events and the KONUS events, it is likely that the BATSE events originate from SGR1900+14.

The fluences of the four pulses lie in the range $(4.5\text{--}6.6) \times 10^{-8}$ erg cm^{-2} (Table 1). These values are ~ 5 times weaker than

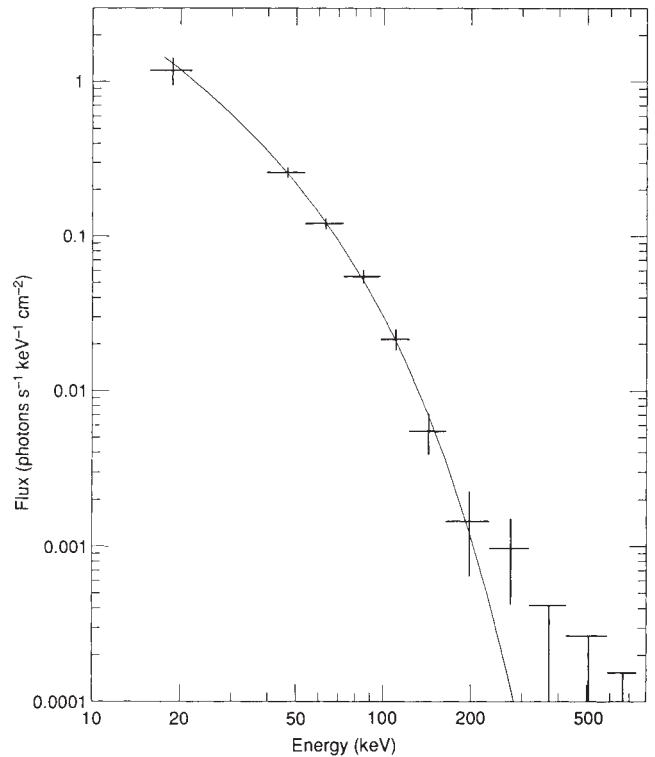


FIG. 2 Photon spectrum of the pulse of Fig. 1b. The spectrum is integrated over 32 ms and corresponds to the peak of the pulse. The gap in the spectrum between ~ 20 and 40 keV is due to a period of telemetry failure during transmission of the data. The curve indicates the best-fit OTTB function, of the form $A \exp(-E/kT)/E$, with $kT = 39 \pm 3$ keV.

those of the weakest event from SGR1900+14 previously reported⁸.

A search of the 1-s-resolution BATSE database for additional, untriggered events shows that during the intervals between the SGR triggers there is an excess of short (< 1 s), low-energy events from the general region of the sky near SGR1900+14 (ref. 17). This indicates that the three triggered events could be accompanied by a much larger number of weaker (and softer) events.

It is possible that SGR1900+14 has had other similar intervals of activity since its discovery in 1979. The sensitivity of previous instruments to these events is difficult for us to estimate, as it depends strongly on the detector energy response, sky coverage and time resolution. The low fluences, soft spectra and short durations of the BATSE events would have made their detection by most previous burst instruments unlikely but not impossible. The fact that we have not yet found other triggered events from the SGR in the BATSE data indicates that its activity is sporadic, but this is not conclusive; a search for untriggered events outside the 'active' period is in progress¹⁷.

The newly discovered Aquila X-ray transient GRS1915+105 (refs 18, 19) is located at $\alpha = 288.2^\circ$, $\delta = 10.9^\circ$, and thus included in the BATSE error box. Inspection of the BATSE flux history of the transient source indicates that it started rising one day

TABLE 1 Temporal and spectral properties of SGR1900+14

Date	UT (s)	RA ($^\circ$)	Dec ($^\circ$)	Duration (ms)	Hardness ratios		Fluence (10^{-8} erg cm^{-2})
					R_{21}	R_{32}	
1992 June 19a	64,668	288	11	40	0.94 ± 0.02	0.16 ± 0.010	4.5 ± 0.03
1992 June 19b				40	0.64 ± 0.01	0.15 ± 0.010	4.3 ± 0.06
1992 July 08	18,957	286	10	< 64	0.72 ± 0.02	0.14 ± 0.010	3.0 ± 0.04
1992 August 19	17,859	291	13	80	0.64 ± 0.01	0.13 ± 0.004	6.6 ± 0.08

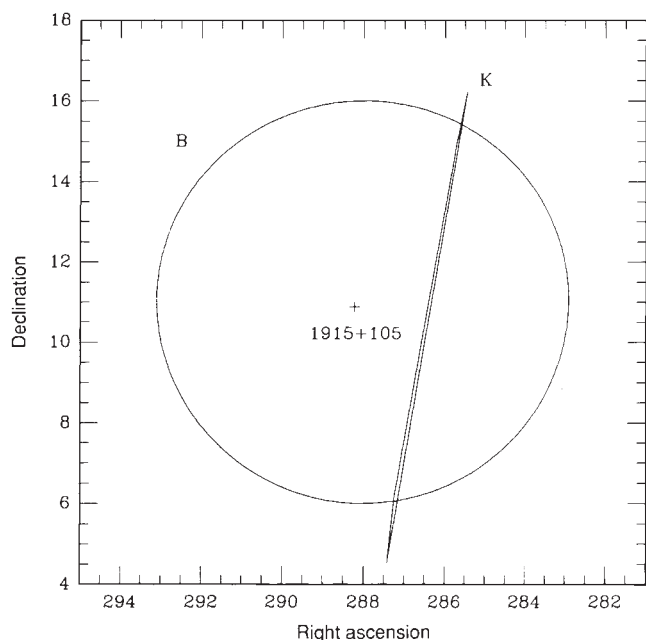


FIG. 3 Sky-map of the region near $\alpha = 288^\circ$, $\delta = +11^\circ$. The large circle (B) represents the average location of the three BATSE events. The size of this circle ($\sim 5^\circ$) reflects mainly the statistical errors of the three measured positions. Burst locations are obtained from the relative strength of their signals in the relevant subset of the eight identical Large Area Detectors (LADs). The position determination includes the detailed spectral dependence of the angular sensitivity of the detectors, and corrections for photon back-scattering by the Earth's atmosphere²³. The diamond (K) represents the 2σ error box of SGR1900+14 (E. Mazets, personal communication). The small cross shows the location of the recently discovered^{18,19} transient X-ray source GRS1915+105 (accuracy $\sim 3'$).

before the first SGR trigger and remained at maximum during the subsequent two triggers. Are the BATSE events related to the X-ray transient? For both the transient and SGR1900+14 the probability of being located inside the BATSE error box by chance is fairly small ($\sim 0.2\%$ for assumed uniform sky distribution). As the positions for both SGR1900+14 and the transient are known with high accuracy (of the order of arcminutes) and they differ by more than $\sim 1.5^\circ$, the KONUS events cannot be related to the Aquila transient. Moreover, a recent analysis²⁰ indicates that there was no persistent X-ray emission (upper limit of $\sim 100 \mu\text{Jy}$) from this region in the sky, during the 1979 SGR bursts. We conclude that it is improbable that the BATSE events originate from GRS1915+105.

Our results suggest that burst activity from the 'old' SGR1900+14 has been detected again ~ 13 years after its discovery. If our detection is indeed the recurrence of activity from this source, it shows that SGRs keep their ability to be active for many years. The extended duration of SGR activity strengthens the argument that these sources are related to galactic (possibly population I) objects, plausibly neutron stars^{4,5}. Recurrent SGR emissions do not signify a unique (catastrophic) event in the life cycle of the source, as is the case in the cosmological models currently favoured for the classical γ -ray bursts^{21,22}. If on the other hand, the new SGR is not related to the SGR1900+14, the case for the SGRs to be associated with population I objects becomes even stronger than it was before, with four (rather than three) sources following their distribution. This is very different from recent results on classical γ -ray bursts, for which a galactic disc origin is excluded^{6,21,22}. The long-term monitoring capability of BATSE gives hope of obtaining valuable information on the recurrence timescale of SGRs and (combined with other spacecraft) accurate source positions, which may lead to a better understanding of the nature of these objects. $\square$

Received 4 January; accepted 9 March 1993.

1. Hurley, K. presented at *Taos Gamma-Ray Stars Conf.* (1986).
2. Atteia, J.-L. et al. *Astrophys. J.* **320**, L105–110 (1987).
3. Laros, J. P. et al. *Astrophys. J.* **320**, L111–115 (1987).
4. Kouveliotou, C. et al. *Astrophys. J.* **322**, L21–25 (1987).
5. Norris, J. P. et al. *Astrophys. J.* **366**, 240–252 (1991).
6. Meegan, C. A. et al. *Nature* **355**, 143–145 (1992).
7. Fishman, G. J. et al. in *Proc. Gamma Ray Observatory Science Workshop* (ed. Johnson, W. N.) 39–50 (NASA/GSFC, Greenbelt, 1989).
8. Mazets, E. P., Golenetskii, S. V. & Guryan, Yu. A. *Sov. Astr. Lett.* **5**(6), 343–344 (1979).
9. Mazets, E. P. et al. *Astr. Space Sci.* **80**, 1–143 (1981).
10. Cline, T. L. *Ann. N.Y. Acad. Sci.* **375**, 314–329 (1981).
11. Lewin, W. H. G., van Paradijs, J. & Taam, R. E. in *X-Ray Binaries* (eds Lewin, W. H. G., van Paradijs, J. & van den Heuvel, E. P. J.) (in the press).
12. Schaefer, B. E. et al. *Astrophys. J.* **393**, L51–L54 (1992).
13. Band, D. et al. *Astrophys. J.* (submitted).
14. Golenetskii, S. V. et al. *Nature* **307**, 41–42 (1984).
15. Kouveliotou, C. et al. in *The Compton Observatory Science Workshop*, NASA CP-3137 (eds Shrader, C. R., Gehrels, N. & Dennis, B.) 61–68 (1992).
16. Lingener, R. E. & Higdon, J. C. *Astrophys. J.* **397**, 576–578 (1992).
17. van Paradijs, J. et al. in *Proc. Compton Symp., St. Louis* (ed. Gehrels, N.) (in the press).
18. Castro-Tirado, A. J. et al. *IAU Circ. No.* 5590 (1992).
19. Cordier, B. et al. in *Proc. Compton Symposium, St. Louis* (ed. Gehrels, N.) (in the press).
20. Lochner, J. & Whitlock, L. *IAU Circ. No.* 5658 (1992).
21. Paczynski, B. *Acta astr.* **41**, 157–166 (1991).
22. Paczynski, B. *Acta astr.* **41**, 257–267 (1991).
23. Brock, M. N. et al. in *Gamma Ray Bursts: Huntsville, AL, 1991* (eds Paciesas, W. S. & Fishman, G. J.) 383–387 (AIP, New York, 1992).

ACKNOWLEDGEMENTS. J.v.P. thanks the BATSE group at the Marshall Space Flight Center and the University of Alabama in Huntsville for their hospitality. He acknowledges financial support from the Leids Kerkhoven Bosscha Fonds.

Discovery of the candidate Kuiper belt object 1992 QB₁

David Jewitt* & Jane Luu†

* Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA

† Department of Astronomy, 601 Campbell Hall, University of California at Berkeley, Berkeley, California 94720, USA

THE apparent emptiness of the outer Solar System has been a long-standing puzzle for astronomers, as it contrasts markedly with the abundance of asteroids and short-period comets found closer to the Sun. One explanation for this might be that the orbits of distant objects are intrinsically short-lived, perhaps owing to the gravitational influence of the giant planets. Another possibility is that such objects are very faint, and thus they might easily go undetected. An early survey¹ designed to detect distant objects culminated with the discovery of Pluto. More recently, similar surveys yielded the comet-like objects 2060 Chiron² and 5145 Pholus³ beyond the orbit of Saturn. Here we report the discovery of a new object, 1992 QB₁, moving beyond the orbit of Neptune. We suggest that this may represent the first detection of a member of the Kuiper belt^{4,5}, the hypothesized population of objects beyond Neptune and a possible source of the short-period comets^{6–8}.

Our observations are part of a deep-imaging survey⁹ of the ecliptic, made with the University of Hawaii 2.2-m telescope on Mauna Kea. The survey uses Tektronix 1,024 × 1,024 pixel and 2,048 × 2,048 pixel charge-coupled devices (CCDs) at the $f/10$ Cassegrain focus. Both CCDs have anti-reflection coatings which yield quantum efficiencies of $\sim 90\%$ at wavelength $\lambda \approx 7,000 \text{ \AA}$ (K. Jim, personal communication). Survey observations are obtained in sets of four images per field with a total timebase of 2 or more hours. Each image is exposed for 900 s while autoguiding at sidereal rate. Because objects in the outer Solar System have small proper motions, our survey was optimized to detect slowly moving objects (SMOs). The angular motions of SMOs are sufficiently small that little trailing-loss results from sidereal tracking. This strategy is found to provide optimum sensitivity to the linear, correlated motion expected of slowly moving objects. By restricting observations to stellar images of full width at half maximum (FWHM) ≤ 1.0 arcsec, and to moonless skies, we obtain limiting magnitudes $m_R \sim 25$. To date, a

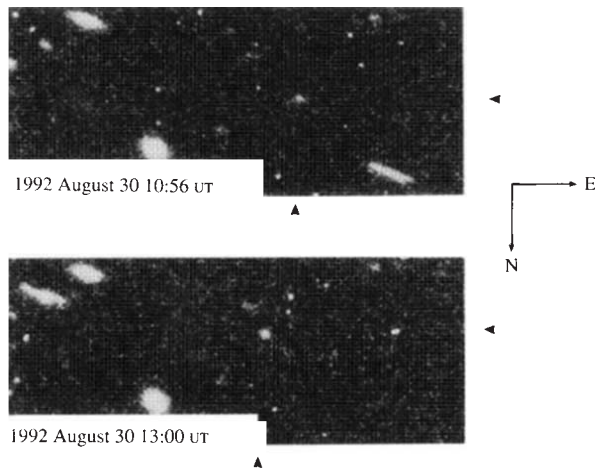


FIG. 1 Discovery images of 1992 QB₁ (marked by arrows), obtained UT 1992 August 30 at the University of Hawaii 2.2-m telescope, using a Mould R filter (central wavelength $\lambda_c = 6,500 \text{ \AA}$, FWHM $\Delta\lambda = 1,250 \text{ \AA}$). The images show regions of a $2,048 \times 2,048$ pixel Tektronix CCD (each pixel subtends 0.22 arcsec). Stellar images are about 0.8 arcsec FWHM. The elongated object to the lower right in the top image is a main-belt asteroid; it appears in the top left of the bottom image and demonstrates the extraordinarily slow motion of 1992 QB₁. The field shown is $\sim 90 \text{ arcsec}$ in width.

sky area of 0.7 square degrees has been imaged to this depth. We are extending our coverage to 1 square degree.

We confine our observations to the opposition direction, where the angular motion of distant objects is retrograde and primarily due to the Earth's motion. At opposition, the parallactic angular motion $\dot{\theta}$ (arcsec per hour) is given by $\dot{\theta} = 148[(1 - R^{-1/2})/(R - 1)]$, where R (in astronomical units, AU) is the heliocentric distance⁹. Thus, a measurement of $\dot{\theta}$ yields R directly. Our observations are restricted to the spring and autumnal equinoxes to benefit further from the large galactic latitude of the opposition point, and from the resultant low density of background stars. In fact, field galaxies pose a worse contamination problem than do field stars in our observations.

The object 1992 QB₁ was detected in real time on UT 1992 August 30 (ref. 10). The discovery images are shown in Fig. 1, where the faster motion of a nearby, unnumbered main-belt asteroid emphasizes the remarkably small angular motion (and hence large distance) of QB₁. The slow motion of the SMO ($2.6 \text{ arcsec h}^{-1}$ west and $1.1 \text{ arcsec h}^{-1}$ south) was confirmed on UT August 31 and September 01. Absence of detectable diurnal parallax confirmed that the object was not a near-Earth asteroid whose proper motion fortuitously resembled that of a SMO. Positions of 1992 QB₁ were obtained with reference to stars in the Hubble Guide Star Catalogue. Measurements relative to these stars suggested an absolute astrometric accuracy of $\pm 1 \text{ arcsec}$, and the relative positions were precise to about $\pm 0.3 \text{ arcsec}$. Astrometric positions were communicated to B. Marsden at the Center for Astrophysics. Follow-up astrometry from the 2.2-m telescope was obtained on UT 1992 September 25, and other observers have reported astrometric measurements in the following three months¹¹⁻¹³. Observations from 1992 August 30 to December 25 yielded a current heliocentric distance $R \approx 41 \text{ AU}$, and were used to calculate the orbital parameters listed in Table 1 (ref. 13).

At distance $R = 41 \text{ AU}$, geocentric distance $\Delta = 40 \text{ AU}$ and phase angle $\alpha = 0.5^\circ$, the apparent red magnitude $m_R = 22.8 \pm 0.2$ on August 30 corresponds to absolute magnitude $H_R = 6.6 \pm 0.2$. For comparison, the absolute magnitude of the distant comet 2060 Chiron in its faint state is $H_R \approx 6.3 \pm 0.1$ (ref. 14). Lacking a measurement of the albedo, we cannot determine the size of 1992 QB₁. However, an albedo $p_R = 0.04$, similar to that of a comet nucleus, suggests a diameter $d \approx 250 \text{ km}$, roughly one-eighth the size of Pluto. The Kron-Cousins colours of 1992 QB₁

TABLE 1 Preliminary orbital parameters

Semi-major axis	a	44.4 AU
Eccentricity	e	0.11
Inclination	i	2.2°
Orbital period	P	296 years
Perihelion date	T	AD 2023
Perihelion distance	q	39.6
Aphelion distance	Q	49.1

Based on astrometry in the interval 1992 August 30 to December 25. Orbit solution is by Marsden¹³.

were measured on UT 1992 August 30 and September 01. Our best estimates are $m_V - m_R = 0.6 \pm 0.1$ and $m_R - m_I = 1.0 \pm 0.2$, to be compared with solar colours $m_V - m_R = 0.32$ and $m_R - m_I = 0.4$, respectively. Here m_V , m_R and m_I are the stellar magnitudes measured with V ($5,500 \text{ \AA}$), R ($6,500 \text{ \AA}$) and I ($8,000 \text{ \AA}$) filters. The $m_R - m_I$ colour is based on a single I filter image. Nevertheless, it seems that 1992 QB₁ is substantially redder than sunlight, inconsistent with a surface of pure ice but consistent with dirty ice, or one contaminated with organic compounds. For comparison, the corresponding colours of the distant object 5145 Pholus are $m_V - m_R \approx 0.7$ and $m_R - m_I \approx 0.7$ (ref. 15), and all other known cometary nuclei are substantially less red¹⁶. Both Pholus and 1992 QB₁ may retain primitive organic mantles produced by prolonged cosmic-ray irradiation. Such mantles would be disrupted or buried by rubble mantles on active, near-Sun comets^{15,16}.

Figure 2 shows the surface brightness profile measured from images taken UT 1992 September 01. The profile of a nearby field star is shown for comparison. Figure 2 supplies no evidence for a resolved coma down to surface brightness $29 \text{ mag per square arcsec}$, at a distance of 1.5 arcsec from the centre of the image. The absence of resolved coma allows us to place a model-dependent limit on the mass loss rate from 1992 QB₁, using a profile-fitting method¹⁷. This limit is $dm/dt < 0.7 \text{ kg s}^{-1}$, so 1992 QB₁ is at least 10^4 times less active than a Halley-class, near-Sun comet and probably less active than 2060 Chiron ($dm/dt \approx 1 \text{ kg s}^{-1}$)¹⁸. Weaker activity cannot be constrained by existing observations, however.

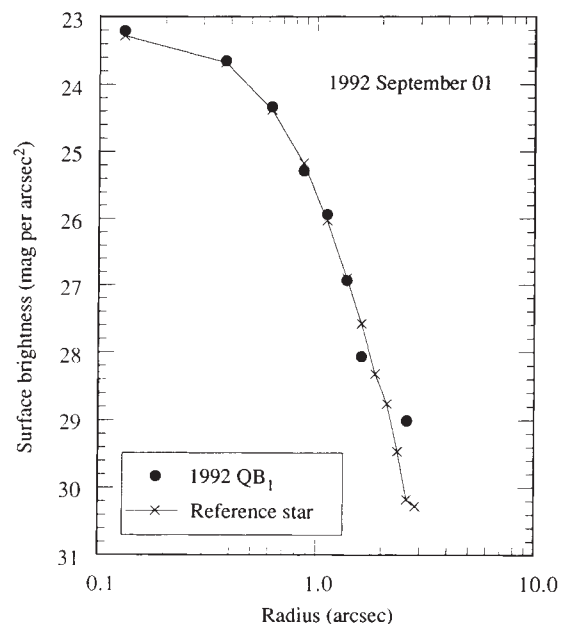


FIG. 2 Surface brightness profile of 1992 QB₁ measured UT 1992 September 01 with 0.8 arcsec FWHM images. The profile was computed from four separate R-filter images of total integration time $3,600 \text{ s}$. A scaled profile of a nearby field star is shown for reference. The image of 1992 QB₁ is consistent with a point source down to surface brightness ~ 29 magnitudes per square arcsec.

As our survey is still in progress, we will not discuss in detail the statistics implied by the detection of 1992 QB₁. It is, however, interesting to make a crude estimate of the Kuiper belt population. We note that 1992 QB₁ was detected in a survey that has so far covered 0.7 square degree. The implied surface density of similar objects is of order 1 per square degree. The inclination of 1992 QB₁ is 2 degrees (Table 1), so the angular width of the belt is at least 4 degrees, and is presumably much larger. A lower limit to the projected area of the Kuiper belt is $360 \times 4 = 1,440$ square degree, and the number of similar objects is thus $N \geq 1,440$. A minimal belt mass, based on an assumed diameter of 250 km and density of 10^3 kg m^{-3} for each object, is $M \geq 1 \times 10^{22} \text{ kg} \approx 2 \times 10^{-3} M_{\text{Earth}}$ (about 1 Pluto mass). This is compatible with the upper limit to the belt mass inferred from dynamical observations of P/Halley, $M \approx 1 M_{\text{Earth}}$ (refs 19–21), and with the minimal mass needed to supply the observed flux of short period comets, $M \approx 0.02 M_{\text{Earth}}$ (ref. 6). The estimated mass must be augmented to account for Kuiper belt comets too small or too distant to be detected in the present survey, but this depends on the (poorly known) mass and spatial distribution of cometary nuclei. M is a lower limit to the true mass.

The faintness of 1992 QB₁ makes it unlikely that earlier observations will be identified, so limiting the accuracy of the derived orbital parameters. The astrometric timebase used to obtain the parameters in Table 1 is only 10^{-3} of the orbit period. Although these parameters are not finalized, it is likely that the perihelion lies beyond the orbits of the gas giant planets (the expected perihelion $q \approx 40 \text{ AU}$, compared with the Neptune aphelion $Q_N = 30.3 \text{ AU}$)¹³. It thus seems that 1992 QB₁ escapes strong planetary perturbations and is, in this sense, more primitive than its presumed cousins at smaller distances, 2060 Chiron and 5145 Pholus. Theory suggests that, although many orbits between the giant planets are unstable on timescales short compared with the age of the Solar System, orbits beyond Neptune may be stable for longer times^{22–27}. In particular, low-eccentricity orbits with semi-major axes of $\sim 44 \text{ AU}$ (Table 1) have lifetimes in excess of 10^9 yr (ref. 27), supporting the idea that 1992 QB₁ is a Kuiper belt comet.

Note added in proof: On UT 1993 March 28 we detected a second slow moving object (IAU Circ. 5370, 29 March 1993). This object, 1993 FW, has the same magnitude and angular motion as 1992 QB₁, and is presumably also resident in the Kuiper belt. Our detection of 1993 FW supports the Kuiper belt population estimate given above. □

Received 13 January; accepted 30 March 1993.

1. Tombaugh, C. W. in *Planets and Satellites* (eds Kuiper G. P. & Middlehurst, B. M.) 12–30 (Univ. Chicago Press, Chicago, 1961).
2. Kowal, C., Liller, W. & Marsden, B. in *Dynamics of the Solar System. Proc. IAU Symp. No. 81* (ed. Duncombe R.) 245–249 (Reidel, Dordrecht, 1979).
3. Scottie, J. V. *IAU Circ. No. 5434* (1992).
4. Edgeworth, K. E. *Mon. Not. R. Astr. Soc.* **109**, 600–609 (1949).
5. Kuiper, G. P. in *Astrophysics* (ed. Hynek, J. A.) 357–424 (New York, McGraw-Hill, 1951).
6. Duncan, M., Quinn, T. & Tremaine, S. *Astrophys. J.* **328**, L69–L73 (1988).
7. Whipple, F. *Proc. Nat. Acad. Sci.* **51**, 711–717 (1964).
8. Fernandez, J. A. *Mon. Not. R. Astr. Soc.* **192**, 481–491 (1980).
9. Luu, J. X. & Jewitt, D. C. *Astr. J.* **95**, 1256–1262 (1988).
10. Jewitt, D. C. & Luu, J. X. *IAU Circ. No. 5611* (1992).
11. McNaught, R. & Steel, D. *Minor Planet Circ.* 20878 (1992).
12. Hainaut, O. & Elst, E. *Minor Planet Circ.* 20902 (1992).
13. Marsden, B. G. *IAU Circ. No. 5684* (1992).
14. Luu, J. X. & Jewitt, D. C. *Astr. J.* **100**, 913–932 (1990).
15. Mueller, B., Tholen, D., Hartmann, W. & Cruikshank, D. *Icarus* **97**, 150–154 (1992).
16. Jewitt, D. C. *Cometary Photometry in Comets in the Post-Halley Era* (ed. Newburn, R.) 19–65 (Kluwer, Dordrecht, 1990).
17. Luu, J. X. & Jewitt, D. C. *Icarus* **97**, 276–287 (1992).
18. Luu, J. X. & Jewitt, D. C. *Astr. J.* **100**, 913–932 (1990).
19. Hamid, S. E., Marsden, B. & Whipple, F. *Astr. J.* **73**, 727–729 (1968).
20. Yeomans, D. K. *Proc. 20th ESLAB Symp.* 419–425 (ESA SP-250, Heidelberg, 1986).
21. Hogg, D. W., Quinlan, G. D. & Tremaine, S. *Astr. J.* **101**, 2274–2286 (1991).
22. Torbett, M. V. *Astr. J.* **98**, 1477–1481 (1989).
23. Torbett, M. V. & Smoluchowski, R. *Nature* **345**, 49–51 (1990).
24. Quinn, T., Tremaine, S. & Duncan, M. *Astrophys. J.* **355**, 667–679 (1990).
25. Duncan, M. & Quinn, T. in *Protostars and Planets III* (Tucson, in the press).
26. Holman, M. J. & Wisdom, J. *Astr. J.* (in the press).
27. Levison, H. & Duncan, M. *Astrophys. J.* (in the press).

ACKNOWLEDGEMENTS. We thank B. Marsden for discussions about 1992 QB₁, G. Luppino for CCD cameras, the Institute for Astronomy TAC for consistent allocation of time to this project, and A. Pickles for obtaining supporting astrometric images. The Planetary Astronomy program of NASA provides financial support at the UH 2.2-m telescope. J.X.L. is in receipt of a Hubble Fellowship.

Dust particle impacts during the Giotto encounter with comet Grigg–Skjellerup

J. A. M. McDonnell*, N. McBride*, R. Beard*, E. Bussoletti†, L. Colangeli‡, P. Eberhardt§, J. G. Firth||, R. Grard¶, S. F. Green*, J. M. Greenberg*, E. Grün**, D. W. Hughes††, H. U. Keller‡‡, J. Kissel**, B. A. Lindblad§§, J.-C. Mandeville|||, C. H. Perry||, K. Rembor¶¶, H. Rickman***, G. H. Schwehm¶¶, R. F. Turner||, M. K. Wallis*** & J. C. Zarnecki*

* Unit for Space Sciences, Physics Laboratory, University of Kent, Canterbury, Kent CT2 7NR, UK

† Istituto di Fisica Sperimentale, Istituto Universitario Navale, Via A. De Gasperi 5, 80133 Napoli, Italy

‡ University of Cassino, Via Zamosch 43, 03043 Cassino (Fr), Italy

§ Physikalisches Institut, Universität Bern, CH-3012 Bern, Sidlerstrasse 5, Switzerland

|| Space Science Department, Rutherford Appleton Laboratory, Chilton, Didcot OX11 0QX, UK

¶ PSS, ESTEC, Postbus 299, 2200 AG Noordwijk, Netherlands

** Rijksuniversiteit te Leiden, Huygens Laboratorium, Postbus 9504, 2300 RA Leiden, Netherlands

*** Max-Planck-Institut für Kernphysik, D-6900 Heidelberg, Germany

†† Department of Physics, University of Sheffield, Sheffield S3 7RH, UK

‡‡ Max Planck Institut für Aeronomie, D-3411 Katlenburg-Lindau, Germany

§§ Lund Observatory, Box 43, S-221 00 Lund, Sweden

||| ONERA-CERT-DERTS, BP No. 4025 31055 Toulouse Cedex, France

¶¶ Astronomical Observatory, Box 515, S-751 20 Uppsala, Sweden

¶¶ Fakultät für Physik, Universität Karlsruhe, D-7500 Karlsruhe, Germany

*** School of Mathematics, University of Wales, Senghennydd Road, Cardiff CF2 4AG, UK

IN the European Space Agency's 1992 Giotto Extended Mission, the Dust Impact Detection System operated successfully during a fly-by that took the spacecraft within about 200 km of the nucleus of comet Grigg–Skjellerup. During the encounter, three meteoroid impacts were detected on Giotto's front shield. The particle masses were found to be $100^{+105}_{-50} \mu\text{g}$, $2^{+4}_{-1} \mu\text{g}$ and $20^{+25}_{-10} \mu\text{g}$, suggesting that the mass distribution of the cometary dust was dominated by larger particles. This is supported by the independent detection of a very large meteoroid (14^{+40}_{-4} mg) by the Giotto Radio-Science Experiment, and is consistent with data over the same mass range from the 1986 encounter with comet Halley. The results indicate a higher rate of mass loss from the nucleus than previously thought, and hence a higher dust-to-gas mass ratio.

The Dust Impact Detection System (DIDSY) was designed to measure the flux of dust particles in the mass range 10^{-19} kg to $>10^{-6} \text{ kg}$ in the coma of comet Halley (Fig. 1, refs 1 and 2). The DID1 IMP-P sensor suffered impact damage at the Halley encounter, and at Grigg–Skjellerup, the instrument was noisy and gave no interpretable results. The DID1 IMP-M sensor, however, was fully operational. The DID7 CIS sensor also suffered impact degradation at Halley, and was largely non-operational. The piezoelectric momentum sensors (DID2 to 5) were fully operational.

The 1986 Halley encounter geometry was such that particles struck the front shield at normal incidence with relative velocity 68.4 km s^{-1} . A hypervelocity particle having momentum mv will transfer momentum ϵmv to the shield (the enhancement is due to target ejecta) where ϵ is the momentum enhancement factor. The value taken at Halley³ was $\epsilon = 11$. At Grigg–Skjellerup the trajectories of particles striking the shield made a 21.2° angle with the shield at a relative velocity of 13.8 km s^{-1} . These factors reduced ϵ considerably. Experiments using the 2-MV dust accelerator at the Unit for Space Sciences, Kent, yield a value of $\epsilon = 3.1 \pm 1.0$. This reduced value of ϵ raises the piezoelectric

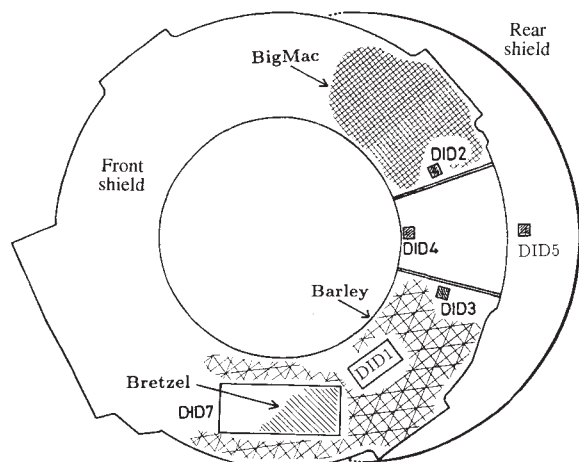


FIG. 1 Giotto is protected from hypervelocity dust impacts by two annular meteoroid bumper shields. The front shield consists of 1-mm-thick aluminium and is situated 250 mm above a rear Kevlar shield. Impacts are detected by three separate subsystems. The first comprises four piezoelectric sensors. Three of these sensors (DID2, DID3 and DID4) are mounted on the rear of the front shield and are sensitive to the momentum imparted to the shield. DID2 and DID3 can operate independently, or measure in coincidence. DID4 operates independently and is located on a 30° section of the shield isolated by acoustically insulating joints. Another piezoelectric sensor (DID5) is situated on the rear shield behind the DID4 section. The impact plasma and momentum subsystem (IPM, DID1) consists of two subunits, located behind an aperture in the front shield. The plasma subunit (IPM-P) determines particle mass by measuring the plasma produced when a particle strikes a metal target plate. The momentum subunit (IPM-M) uses a piezoelectric detector situated on the rear of the metal target plate. IPM-P and IPM-M can measure in coincidence. The capacitor impact sensor subsystem (CIS, DID7) is a thin-film, dielectric-filled capacitor fixed to the front shield. Impacts are indicated by brief discharging. The shaded areas in the figure indicate the possible impact sites of the three meteoroids detected at the Grigg-Skjellerup encounter by DIDSY: Big Mac, Barley and Bretzel.

mass detection threshold. Also, the reduced resolved area of the front shield, and the fact that the activity of Grigg-Skjellerup was estimated to be ~1% that of Halley, meant that few detections were expected. Pre-encounter modelling indicated total fluences between 0 and 40 particles for miss distances ranging between 300 and 50 km respectively. The data gathered were wholly consistent with these predictions. DIDSY was operated for ~35 hours from 18:42 on 9 July, with an impact time

FIG. 2 Cumulative fluence for the three impacts detected by DIDSY (circles). The cumulative mass index is $\alpha = -[0.27^{+0.13}_{-0.20}]$ (1 sigma errors; see Box 1). The steepest (3 sigma) slope consistent with the three detections is $\alpha = -0.74$. The dotted line indicates upper mass cut-off at a fluence equivalent to one particle per resolved spacecraft area. The fluence corresponding to the particle responsible for the spacecraft nutation detected by GRE (square) has not been included in the DIDSY regression data set. The mass error bars plotted are the standard deviations of the mass probability functions and represent uncertainties in impact position. The systematic errors mean that the mass scale could move either way by 0.3, as indicated by the arrows near the bottom of the figure (the GRE particle would remain fixed with respect to the axis). Particle diameters in this figure (and Fig. 3) are calculated assuming a density of 1 g cm^{-3} .

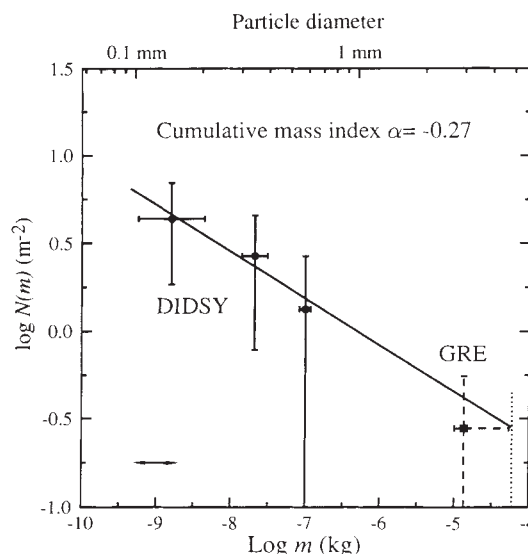
Event	Name	Time (GRT)	Mass
Expected closest approach ⁵		15:30:36 ± 14 s (1σ)	
Nutation particle GRE ⁵	Whopper	15:30:43	14^{+40}_{-4} mg
Maximum optical brightness OPE ⁴		15:30:43 ± 3 s	
DIDSY event 1, DID2 and DID4	Big Mac	15:30:48–51	$100^{+105}_{-60} \mu\text{g}$
DIDSY event 2, DID3	Barley	15:30:51–54	$2^{+4}_{-1} \mu\text{g}$
DIDSY event 3, DID3 and CIS	Bretzel	15:31:31–32	$20^{+25}_{-10} \mu\text{g}$

The errors quoted (to 1 sigma) on the three DIDSY particle masses incorporate both the statistical and systematic errors and represent the absolute uncertainty in the particle masses.

resolution of 2.83 s. All times given here are ground receive times in UT (GRT = spacecraft time + 11 m 53 s).

Three particle impacts were detected, by sensors DID2, DID3 and DID4, compared with 2,400 at the Halley encounter. Because of the impact degradation of CIS, the bias voltage was 0.22 V as opposed to 2.3 V at Halley. This voltage was stable for 21 h up to closest approach, changed to 0.20 V at 15:31:31–32 and then remained stable for the rest of the operating period. This change coincided with the impact interval of the third particle, and indicates an impact directly on CIS. The CIS voltage was monitored roughly once a second giving a better time resolution on event 3. The times and the masses of the three meteoroids are given in Table 1 (see Box 1 for explanation of mass determination). The first two events were near closest approach when the spacecraft-comet distance⁴ was likely to have been <200 km and the third event was in excess of 40 s after closest approach at >500 km from the nucleus.

The wide distribution of masses yielded a most significant and unexpected mass distribution. The fluence, being the time integrated particle flux, is shown in Fig. 2. A most probable upper mass cut-off at $10^{-4.23} \text{ kg}$ (59 mg) was calculated for one detected particle hitting the entire resolved spacecraft area (~3.6 m²). The cumulative mass distribution index α was found to have a value of $\alpha = -0.27$. The error in this gradient is dominated by the statistical number uncertainty rather than the particle mass uncertainty. An analysis based on the null detection of a fourth DIDSY particle (see Box 1) indicates that $\alpha = -[0.27^{+0.13}_{-0.20}]$ (1 sigma errors) and could be as high as -0.74 (3 sigma limit).



BOX 1 Particle mass determination

The signal attenuation for an impact on a particular area element of the front shield is known from laser shot mapping done at the Rutherford Appleton Laboratory². A probability distribution of particle masses is obtained by identifying the mass required to produce the observed sensor signal from each area element. In examining all possible impact sites, a cumulative mass distribution $N(m) = km^\alpha$ is assumed so that a probability of the particle being between mass m and $m + dm$ can be calculated thus:

$$P(m) = \left[\sum_{i=1}^{\text{shield}} dN(m) a_i f(m) \right] \times \left[\sum_{m=0}^{\infty} \sum_{i=1}^{\text{shield}} dN(m) a_i f(m) \right]^{-1}$$

where dN is the differential mass distribution function (the number of particles of mass m to $m + dm$) and a_i is the area of the i th shield element. The function $f(m) = 1$ if the mass m gives the observed signal for this area element, and $f(m) = 0$ if it does not. This is an iterative process where particle mass is determined using an initial chosen value of α . The mass distribution is then calculated using the effective shield areas (see below) so yielding a new value of α . The process is repeated until α converges. The determined particle mass is not, however, strongly dependent on α .

There are two sorts of error in the particle mass determination. The first is that described by the standard deviation of the pseudo-gaussian mass probability function due to uncertainties in impact position. The second is the systematic errors due to uncertainty in the absolute sensor sensitivity and momentum enhancement factor. The probability errors affect the determination of α . The systematic errors do not affect the relative positions of the mass probability functions (or the mass threshold, see below) but have the effect of moving the mass scale (Fig. 2) either way by a factor of 2 (0.3 in log space). Hence this error does not affect the determined mass distribution index.

The fluence is determined from the observed number of impacts by particles of mass $\geq m$, and the effective shield area that is sensitive to these particles, which, as we are calculating a cumulative fluence, is also a function of $N(m)$. The geometric area of the front shield is 2.09 m². The effective shield sensing areas are 2.09, 1.92 and 2.09 m² for events 1, 2 and 3 respectively. The cumulative mass distribution index is found to be $\alpha = -0.27$. An uncertainty in α can be determined by considering the effect of tilting the straight-line fit about the error weighted mean in Fig. 2, and considering when the three detections would not be consistent with this gradient, within statistical errors. Figure 3 shows the expected number of detections with mass $\geq m$, calculated for a range of α values. The mass threshold (which is weakly dependent on α) is shown as a line cutting each curve, and is defined as the mass at which 95% of all particles (with this mass or greater) hitting the shield would be detected by the DID2, 3 or 4 sensors. Values of α between -0.07 and -0.40 are consistent with the detection of three particles within 1 sigma statistical errors. A 3-sigma limit corresponds to $\alpha = -0.74$. This statistical uncertainty is dominant over the mass probability error. The systematic errors again do not affect this analysis.

Additional support for a low-slope mass distribution was given by results from the Giotto Radio-Science Experiment⁵ (GRE) which indicated that a particularly large meteoroid hit the rear of the spacecraft (hence no DIDSY detection). The effective mass (where $m_{\text{eff}} = \epsilon m_{\text{real}}$) was calculated to be between 30 and 54 mg. Taking $\epsilon = 3.1$, as before, gave the most likely value of the real mass of the particle as ~ 14 mg with the lower limit being 10 mg. It is possible for a penetrating impact to have little or no momentum enhancement, and so a reasonable upper limit is 54 mg. This particle is also plotted in Fig. 2, although it was not used to determine α , to show that the slope of -0.27 is entirely consistent with the detection of this particle and may hold over four decades of mass. An integration between mass limits $10^{-9.35}$ and $10^{-4.23}$ kg gives the total mass that struck Giotto as ~ 22 mg. With such a low slope, this mass is dominated by

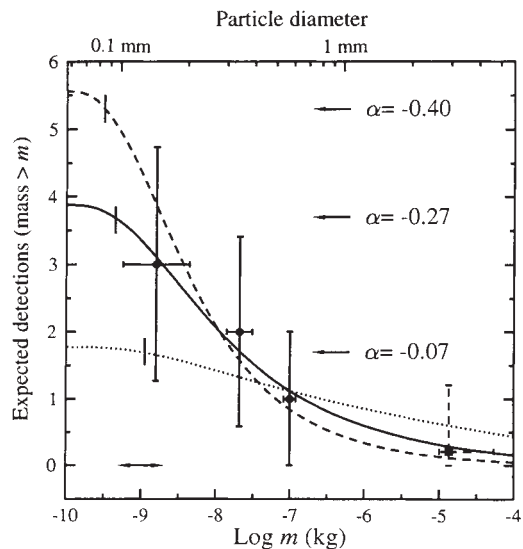


FIG. 3 The expected number of detections of particles with mass $\geq m$. The solid curve is calculated using $\alpha = -0.27$. The dotted and dashed curves show the lower and upper limits within which three DIDSY detections with the mass above the threshold mass are consistent (within 1 sigma statistical limits). These curves are calculated using $\alpha = -0.07$ and $\alpha = -0.40$ respectively. The small solid vertical lines cutting each curve indicate the mass threshold. As in Fig. 2, systematic errors mean that the mass scale could move either way by 0.3, indicated by the arrows near the bottom of the figure (again the GRE particle would remain fixed with respect to the axis).

the largest particle to strike and is consistent with the above meteoroid. Assuming a density of 1 g cm^{-3} , the cumulative cross-sectional area of the particles between these mass limits is 12 mm^2 .

The low slope found from the three DIDSY particles also agrees well with Halley data over the same mass range of 10^{-9} to 10^{-7} kg, where the time-averaged mass distribution index was found to be -0.2 . From 10^{-9} to 10^{-15} kg, the Halley index increased to -0.9 . Although the Grigg-Skjellerup data presented here say nothing about masses $< 10^{-9}$ kg as such, it is likely that the index also increases at smaller mass to account for the presence of optical scattering particles⁶.

DIDSY can give no information as to the composition of impacting particles (solid or aggregate), and the number of detections was not high enough to give information about the variation of the mass distribution index with time. The very low slope for the mass distribution, however, agrees with the Halley data and IRAS dust trail data. The dominance of the large particles is important for modelling nucleus structure and composition, and for modelling the contribution and evolution of cometary particles to the interplanetary dust flux. It also signifies a higher rate of mass loss than previously thought and hence a higher ratio (by mass) of dust to gas. □

Received 2 November 1992; accepted 2 March 1993.

- McDonnell, J. A. M. et al. *ESA spec. Publ.* **1077**, 85–108 (1986).
- McDonnell, J. A. M. *J. Phys. E: Sci. Instrum.* **20**, 741–758 (1987).
- McDonnell, J. A. M. et al. *Adv. Space Res.* **4**, 297–301 (1984).
- Levasseur-Regourd, A. C. et al. *Nature* (submitted).
- Pätzold, M. P., Edenhofer, M. K., Bird & H. Volland, *Astr. Astrophys.* (in the press).

ACKNOWLEDGEMENTS. We acknowledge the UK SERC for support. We thank N. Schmidt for his help during encounter, and especially acknowledge the efforts of G. Schwehm and M. Gensemann.

Hydrogen bonding in the benzene–ammonia dimer

David A. Rodham*, Sakae Suzuki*,
Richard D. Suenram†, Frank J. Lovas‡,
Siddharth Dasgupta‡, William A. Goddard III‡
& Geoffrey A. Blake§

* Division of Chemistry and Chemical Engineering, 127-72,
California Institute of Technology, Pasadena, California 91125, USA

† Molecular Physics Division, National Institute of Standards and
Technology, Gaithersburg, Maryland 20899, USA

‡ Materials and Molecular Simulation Center, Beckman Institute, 139-74,
California Institute of Technology, Pasadena, California 91125, USA

§ Division of Geological and Planetary Sciences, 170-25,
California Institute of Technology, Pasadena, California 91125, USA

AMINES have long been characterized as amphoteric (acting as both donor and acceptor) in terms of their hydrogen-bond interactions in the condensed phase. With the possible exception of $(\text{NH}_3)_2$, however, no gas-phase complexes exhibiting hydrogen-bond donation by ammonia, the 'simplest amine', have been observed^{1,2}. Here we present high-resolution optical and microwave spectra of the benzene–ammonia dimer in the gas phase, which show that the ammonia molecule resides above the benzene plane and undergoes free or nearly free internal rotation. In the vibrationally averaged structure, the C_3 symmetry axis of NH_3 is tilted by about 58° relative to the benzene C_6 axis, such that the ammonia protons interact with the benzene π -cloud. Our *ab initio* calculations predict a 'monodentate' minimum-energy structure, with very low barriers to rotation of ammonia. The larger separation of the two molecular components, and the smaller dissociation energy, relative to the benzene–water dimer³ reflect the weak hydrogen-bond donor capability of ammonia, but the observed geometry greatly resembles the amino–aromatic interaction found naturally in proteins⁴.

Typically, hydrogen-bond interactions have energies of 8–29 kJ mol^{-1} (2–7 kcal mol^{-1}) and are thought to involve essentially linear arrangements of donors, such as $-\text{OH}$ or $-\text{NH}$ groups, with a proton acceptor, often a highly electronegative atom. From this 'classical' point of view, the interactions of hydrogen-bond donors with aromatic systems are of both theoretical and practical interest as they can be used to probe the general limits of hydrogen-bonding functional groups⁵. The amino–aromatic and ammonium–aromatic interactions are common motifs in protein structure and in rational molecular design^{4,6}. Crystallographic imaging of specific interactions, such as that in the binding region of the *v-src* SH2 domain⁷, is consistent with 'hydrogen-bond' donation from the $-\text{NH}_2$ and $-\text{NH}_3^+$ groups, although the *in vivo* stability of such structures is difficult to estimate⁸.

In principle, the energetics and structural parameters of the amino–aromatic interaction, free of external perturbations, may be derived from isolated weakly bound clusters. The simplest of these would be the benzene–ammonia dimer. But gas-phase ammonia can best be described as a powerful proton acceptor¹. In recent work on $\text{C}_6\text{H}_6\text{--H}_2\text{O}$ (ref. 3), we found that benzene does act as a 'hydrogen-bond' acceptor in this complex. Given the similarity of the observed structure to analogous hydroxyl–aromatic and amino–aromatic interactions found in proteins, we started the present study on the benzene–ammonia dimer.

We have used resonance-enhanced two-photon ionization (R2PI) and microwave spectroscopy along with quantum chemical calculations to study the structure and energetics of the $\text{C}_6\text{H}_6\text{--NH}_3$ system. Ionization is enhanced when the energy of the ionizing laser is in resonance with a rovibronic transition of the cluster. The fine structure of the resonances provides insights into the rotational dynamics of the cluster, and thus on the bonding geometry. Typical R2PI spectra through the

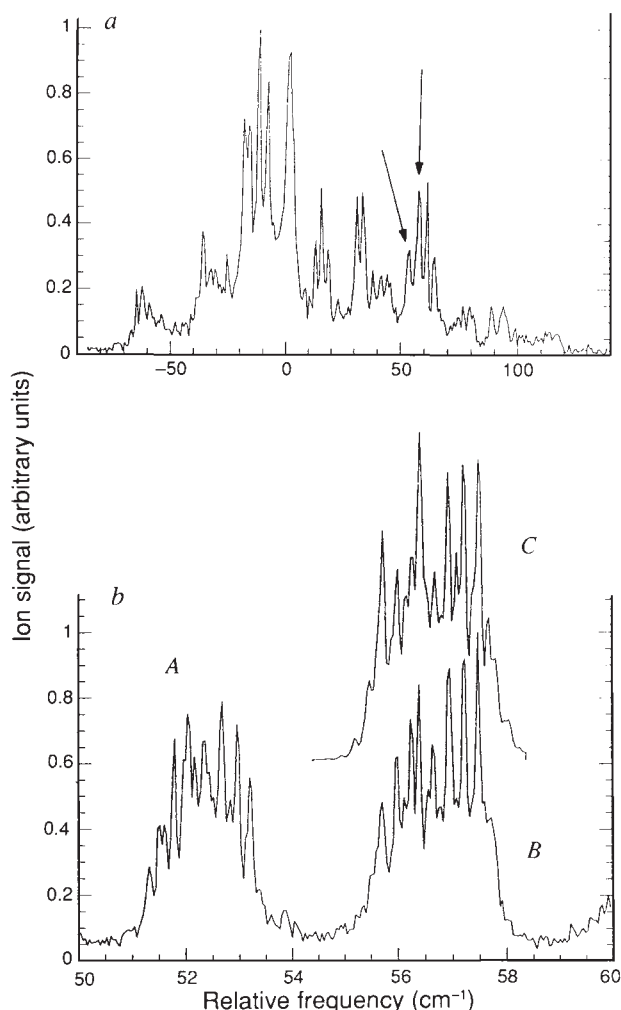


FIG. 1 R2PI spectra of $\text{C}_6\text{H}_6\text{--NH}_3$, obtained by ionizing clusters in a skimmed molecular beam with the frequency-doubled output of an excimer-pumped dye laser in the source region of a reflection time-of-flight mass spectrometer. *a*, Low-resolution spectrum (0.4 cm^{-1} laser bandwidth, 0.75 cm^{-1} per point) recorded by monitoring the $[\text{C}_6\text{H}_6\text{--NH}_3]^+$ mass channel at high NH_3 concentration (0.27%). The frequency scale is relative to the 6_0^1 transition of free benzene. Arrows point to the 6_0^1 transition of the 1:1 complex; most of the other features are fragmentation products of larger clusters. *b*, Higher-resolution (0.08 cm^{-1} laser bandwidth, 0.04 cm^{-1} per point) scan of the 1:1 complex 6_0^1 band. *A* is assigned to an excited $m=1$ internal rotor state, *B* originates from the ground $m=0$ internal rotor state, and *C* is a calculated fit of the $m=0$ band. For this case of a symmetric top with a free internal rotor, the *A* rotational constants for the S_0 and S_1 electronic states of the complex should be close to the *c*-axis constants of free benzene, as is observed. Using the free rotor (benzene *c*-axis) values of $A''=0.0948 \text{ cm}^{-1}$ and $A'=0.0906 \text{ cm}^{-1}$ plus the *B''* ground state rotational constants from Table 1, the best fits produced $B'=0.0640 \text{ cm}^{-1}$, $\zeta'=-0.48$ for $\text{C}_6\text{H}_6\text{--NH}_3$ and $B'=0.0578 \text{ cm}^{-1}$, $\zeta'=-0.58$ for $\text{C}_6\text{H}_6\text{--ND}_3$. The benzene $S_1(6^1)$ Coriolis parameter is $\zeta=-0.58$, and the highest correlation in the fits is between *A'* and ζ .

$S_1 \leftarrow S_0(6_0^1)$ transition of benzene are shown in Fig. 1. Results for ND_3 mixtures were similar. Features assigned to the dimer remained at low ammonia concentrations (0.01%), and were the only peaks for which cluster fragmentation produced significant numbers of free benzene cations. Most of the other spectral features in Fig. 1*a* resulted from the efficient fragmentation of larger clusters after ionization. Earlier R2PI spectra of $\text{C}_6\text{H}_6\text{--NH}_3$ were extremely complex because of such fragmentation⁹, which provides indirect evidence for a hydrogen-bonded geometry¹⁰.

High-resolution R2PI scans (Fig. 1*b*) demonstrate that the

TABLE 1 Spectroscopic fits to $m=0$ microwave data

Complex	B (MHz)	D_J (kHz)	D_{JK} (kHz)	σ (MHz)	
$C_6H_6-^{14}NH_3$	1,889.026 (3)	4.03 (4)	91.52 (9)	0.038	$eQq_{aa}=0.286$ (15) MHz $\mu_a=3.814(15) \times 10^{-1}$ cm (1.142 (5) D)
$C_6H_6-^{15}NH_3$	1,830.4272 (7)	3.87 (3)	89.50 (6)	0.003	
$C_6H_6-^{14}ND_3$	1,732.986 (7)	3.77 (3)	127.79 (51)	0.030	

Spectra of the $J=6 \leftarrow 5$ and $J=7 \leftarrow 6$ transitions of $C_6H_6-^{14}NH_3$ were initially recorded using the Stark modulation microwave absorption spectrometer at the California Institute of Technology. Clusters were formed by expanding Ar seeded with $\sim 1\%$ of both C_6H_6 and NH_3 through a 50 mm $\times$ 0.025 mm slit nozzle. Spectra of the $J=3 \leftarrow 2$ and $4 \leftarrow 3$ transitions of benzene complexed with $^{14}NH_3$, $^{15}NH_3$ and $^{14}ND_3$ were subsequently obtained with the Fourier transform microwave spectrometer at NIST-Gaithersburg. Both instruments have been described in detail previously^{17,18}. Transition frequencies were fitted using the standard symmetric-top equation $\nu(J \rightarrow J-1) = 2J(B - D_{JK}K^2) - 4D_JJ^3$, where B is the rotational constant and D_J and D_{JK} are the 'stretching' and 'bending' centrifugal distortion constants of the dimer. σ is the root-mean-square error of the fit. Dipole moments for the $^{14}NH_3$ and $^{15}NH_3$ complexes were determined from the Stark shifts of the $K=0$ lines in a static electric field, and are the same within experimental error. The $J=3 \leftarrow 2$, $K=2$ transition was used to determine the quadrupole moment of $C_6H_6-^{14}NH_3$. From these measurements, the angular projections were calculated in the standard fashion: $\mu_a(C_6H_6-NH_3) = \mu_{ind} + \mu_{NH_3}(\cos \theta)$; $eQq_{aa}(C_6H_6-NH_3) = eQq_{NH_3}([3 \cos^2 \theta - 1]/2)$. The measured $P_2(\cos \theta)$ and μ_a values produce a μ_{ind} of 1.0×10^{-30} C m (0.3 Debye units, D), which is consistent with estimates based on the ammonia dipole moment and the polarizability of benzene when $\langle \theta \rangle \sim 60^\circ$. The *ab initio* calculations find $\theta^\circ = 52^\circ$, $\mu(NH_3) = 5.67 \times 10^{-30}$ C m (1.89 D), and $\mu_a(C_6H_6-NH_3) = 4.32 \times 10^{-30}$ C m (1.44 D), which yield $\mu_{ind}(MP2/6-31G^{**}) = 1.0 \times 10^{-30}$ C m (0.3 D).

dimer retains symmetry about the benzene C_6 axis on a vibrationally averaged basis. The appearance of two peaks suggests that the ammonia can freely rotate in $C_6H_6-NH_3$, with one band arising from the [$j(NH_3)=0$, $m=0$] ground state and the other from the [$j(NH_3)=1$, $m=1$] internal rotor state. The $m=1$ states are metastable because of nuclear spin eigenfunction restrictions¹⁰. A least-squares fit to the $m=0$ band using nuclear spin statistical weights determined for the G_{36} molecular symmetry group is shown in Fig. 1b (trace C). The G_{36} group considers the internal rotation of the monomers and the exchange of the ammonia protons, but not inversion, to be 'feasible'. The shape of the $m=0$ contour is sensitive to the A rotational constant of the dimer and is consistent only with a symmetric structure in which a nearly freely internally rotating ammonia resides above the benzene plane, suggesting the general structure shown in Fig. 2. Ammonia inversion may well occur in $C_6H_6-NH_3$, but the R2PI data show no direct evidence of it.

These conclusions are strengthened by the symmetric-top pure rotational (microwave) spectra observed for the $m=0$ state (data available from the authors on request). Transitions originating from the $m=1$ states have not yet been observed in the microwave region, but they are not critical to the structural determination outlined below. The rotational and centrifugal distortion constants for three isotopic species ($C_6H_6-^{14}NH_3$, $^{15}NH_3$, $^{14}ND_3$) obtained from the $m=0$ data are listed in Table 1. The rotational constants are consistent with the structure deduced from the R2PI data, as is outlined in Fig. 2. A Lennard-Jones approximation to the radial potential and the centrifugal distortion constant D_J in Table 1 lead to an experimental estimate of ~ 5.9 kJ mol⁻¹ (1.4 kcal mol⁻¹) for the zero-point dissociation energy D_0 and a stretching force constant of 4.1(2) N m⁻¹. The slight decrease and large increase in the values of the distortion constant D_{JK} on isotopic substitution of $^{15}NH_3$ and $^{14}ND_3$ are predicted to within 20% by a harmonic bending potential, which yields a bending force constant of 1–2 N m⁻¹.

This same general structure has been found for the C_6H_6-HCl (ref. 11), C_6H_6-HF (ref. 12), and $C_6H_6-H_2O$ (refs 3, 10) hydrogen-bonded complexes. These dimers, like $C_6H_6-NH_3$, exhibit symmetric-top ground-state spectra as a result of internal rotation or vibrational averaging that maintains symmetry about the benzene C_6 axis. In addition to the rotational constants, the a -axis projection of the nitrogen nuclear quadrupole coupling constant (eQq_{aa}) and electric dipole (μ) moments of the cluster were measured and used to extract NH_3 $P_2(\theta)$ and $P_1(\theta)$ angular projections. The NH_3 tilt angle determined from eQq_{aa} is $\langle \theta^2 \rangle^{1/2} = 57.6 \pm 0.3^\circ$. Uncertainties in the induced dipole (μ_{ind}) moment render the $P_1(\theta)$ value more suspect, but for a range of $\pm 0.33 \times 10^{-30}$ cm (± 0.1 Debye unit (D)) in μ_{ind} we estimate

that $\langle \theta \rangle = 60 \pm 5^\circ$ (see Table 1). The consistency of the angular expectation values, in contrast to those observed for Ar- NH_3 and Ar- H_2O (ref. 13), clearly indicate that the ammonia protons

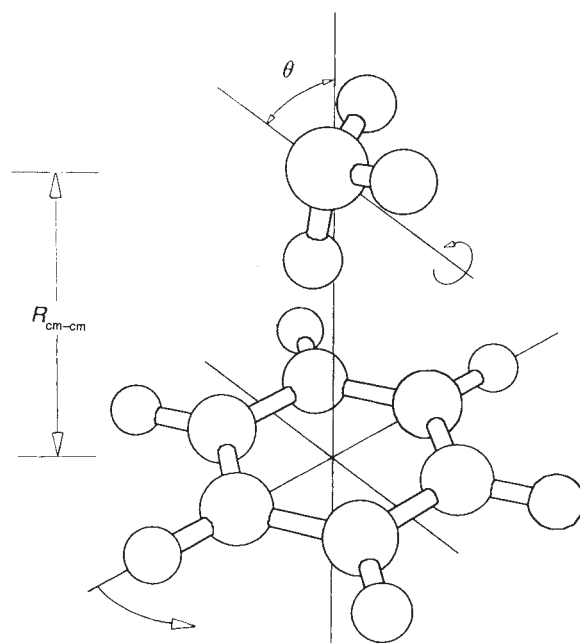


FIG. 2 Coordinate system and structure of the benzene-ammonia dimer. To calculate the weak bond length and the tilt angle of the ammonia, we used the formula¹⁹

$$I_{bb}(C_6H_6-NH_3) = \mu R_{cm-cm}^2 + \frac{I_{bb}(NH_3)}{2}(1 + \cos^2 \theta) + \frac{I_{cc}(NH_3)}{2}(\sin^2 \theta) + I_{bb}(C_6H_6)$$

where θ is the angle between the NH_3 C_3 axis and the benzene C_6 axis, with the ammonia centre of mass on the benzene C_6 axis and $\theta=0$ corresponding to the configuration in which all three ammonia protons point toward the benzene. μ is the pseudodiatom reduced mass of the complex and R_{cm-cm} is the distance between the molecular centres of mass. $I_{cc}(NH_3)$ and $I_{bb}(NH_3)$ are the ammonia moments of inertia parallel and perpendicular to the ammonia C_3 symmetry axis. Arrows depict the feasible large amplitude motions in this complex. Assuming that the distance from the nitrogen atom to the benzene plane R_{N-Bz} was the same for all three isotopic species led to independent constraints on R_{N-Bz} and θ , which were found to be $R_{N-Bz} = 3.590 \pm 0.005$ Å and $\theta = 59 \pm 5^\circ$. The above equation assumes that the vibrational averaging of the benzene orientation is negligible. Tilting of the benzene plane by 5 and 10° reduces the R_{N-Bz} by 0.0035 and 0.015 Å, respectively. Because of the symmetric top character of the monomers, the individual ammonia proton locations cannot be determined experimentally.

are 'bound' to the benzene π -cloud: that is, the anisotropy in the intermolecular potential is substantial.

To estimate the binding energy and other global properties of the intermolecular potential, theoretical calculations were performed for the dimer using the program Gaussian 92 (ref. 14) at the MP2/6-31G** level. Previous *ab initio* efforts on this complex have either neglected electron correlation¹⁵ or not used it in scans of the potential energy surface¹⁶, but show, as here, that the binding well is shallow. The MP2/6-31G** non-BSSE (non-basis set superposition error-corrected) binding energy is $D_e \approx 10 \text{ kJ mol}^{-1}$ (2.4 kcal mol⁻¹). Full energy minimization of the intermolecular degrees of freedom produced an equilibrium structure ($R_{\text{cm-cm}}^e = 3.43 \text{ \AA}$, $\theta^e = 52^\circ$) which is very similar to that experimentally observed, but different from that predicted solely from the leading electrostatic attractive forces (dipole-induced dipole, dipole-quadrupole and so on). The calculations indicate that the geometry where one proton is pointing towards the centre of the benzene ring is slightly more stable than that with two protons pointing towards the benzene ring. In both minima the ammonia centre of mass lies above the plane of benzene along its C_6 axis. The energy difference between the 'monodentate' and 'bidentate' geometry, however, is only 0.4 kJ mol^{-1} (0.1 kcal mol⁻¹), a value comparable to the intermolecular zero-point vibrational energies.

The predicted D_e is at the lowest end of the range commonly accepted for hydrogen bonding—the hydrogen-bond energy in water ice is near 21 kJ mol^{-1} (5 kcal mol⁻¹), and typical O-H...O and N-H...N hydrogen bonds range in strength from

13–25 to 8–21 kJ mol⁻¹ (3–6 to 2–5 kcal mol⁻¹; ref. 5)—and is considerably below that of $C_6H_6 \cdots H_2O$ (ref. 3), consistent with the gas-phase acidities of ammonia and water. Nevertheless, the benzene–ammonia dimer constitutes a rare example of proton donation by ammonia in weakly bound complexes, and provides the simplest molecular model of the biologically important amino–aromatic interaction. □

Received 11 January; accepted 9 March 1993.

1. Nelson, D. D. Jr, Fraser, G. T. & Klemperer, W. *Science* **238**, 1670–1674 (1987).
2. Loeser, J. G. et al. *J. chem. Phys.* **97**, 4727–4749 (1992).
3. Suzuki, S. et al. *Science* **257**, 942–945 (1992).
4. Burley, S. K. & Petsko, G. A. *FEBS Lett.* **203**, 139–143 (1986).
5. Pimental, G. C. & McClellan, A. L. *The Hydrogen Bond* 202–203 (Freeman, San Francisco, 1960).
6. Dougherty, D. & Stauffer, D. A. *Science* **250**, 1558–1560 (1990).
7. Waksman, G. et al. *Nature* **358**, 646–653 (1992).
8. Perutz, M. F. in *The Chemical Bond* (ed. Zewail, A.) 17–30 (Academic, New York, 1992).
9. Wana, J., Menapace, J. A. & Bernstein, E. R. *J. chem. Phys.* **85**, 1795–1805 (1986).
10. Gotch, A. J. & Zwier, T. S. *J. chem. Phys.* **96**, 3388–3401 (1992).
11. Read, W. G., Campbell, E. J. & Henderson, G. *J. chem. Phys.* **78**, 3501–3508.
12. Baiocchi, F. A., Williams, J. H. & Klemperer, W. *J. phys. Chem.* **87**, 2079–2084 (1983).
13. Cohen, R. C. & Saykally, R. J. *J. phys. Chem.* **96**, 1024–1040 (1992).
14. Frisch, M. J. et al. *Gaussian 92* (Gaussian, Pittsburgh, PA, 1992).
15. Cheney, B. V., Schulz, M. W., Cheney, J. & Richards, W. G. *J. Am. chem. Soc.* **110**, 4195–4198 (1988).
16. Bredas, J. L. & Street, G. B. *J. chem. Phys.* **90**, 7291–7299 (1989).
17. Bumgarner, R. E. & Blake, G. A. *Chem. Phys. Lett.* **161**, 308–314 (1989).
18. Suenram, R. D. et al. *J. molec. Spectrosc.* **137**, 127–137 (1989).
19. Fraser, G. T., Lovas, F. J., Suenram, R. D., Nelson, D. D. & Klemperer, W. *J. chem. Phys.* **84**, 5983–5988 (1986).

ACKNOWLEDGEMENTS. We gratefully acknowledge support from the Caltech Beckman Institute (to W.A.G.), NSF, NASA and the David and Lucille Packard Foundation (to G.A.B.).

Seasonal and diel variation in the open ocean concentration of marine snow aggregates

R. S. Lampitt, W. R. Hillier & P. G. Challenor

Institute of Oceanographic Sciences Deacon Laboratory, Wormley, Godalming, Surrey GU8 5UB, UK

MARINE snow, generally defined as aggregated particles of diameter greater than 0.5 mm, is thought to play an important role in oceanic biogeochemical cycles¹. Recent studies have focused on its unusual physical, biological and chemical properties but its temporal variability has received scant attention^{2–6}. Here we report observations of the abundance, volume concentration and size distribution of marine snow over a five-month period at a single site in the Northeast Atlantic. At a depth of 270 m, marine snow particles demonstrated strong seasonal and diel variability. Volume concentrations in spring were about 20 times those in summer and autumn with late morning concentrations up to three times higher than at other times of the day. Our results suggest that the marine snow forms as a result of highly dynamic interactions in the particle pool. We believe that the mid-water biota and their migratory behaviour are responsible for the diel variability; they are therefore likely to have a significant influence on marine snow concentrations and hence on open-ocean material flux.

A camera system was deployed on a mooring above the Porcupine abyssal plain in the mesotrophic Northeast Atlantic (47° 45' N, 19° 28' W) on 20 April 1990 and recovered 154 days later after taking images of a 20-litre volume every 512 minutes. The apparatus was at an average depth of 270 m. After image processing and data analyses, a strong seasonal signal became apparent. In addition, contrary to expectations, both abundance and volume concentration of marine snow tended to be higher in the day than at night (Fig. 1). Two main abundance and concentration peaks occurred (mid-April and mid to late May).

The initial rapid increases in volume concentration were mainly because of increased abundance of the larger particles (>2.5 mm diameter) but later within the peaks, the volume was dominated by the smaller size categories (1.6–2.5 mm) (Fig. 2). The smallest size categories vary only slowly during the year, presumably reflecting their longer residence time in the water column.

Large changes in the particle pool sometimes occurred over short time periods (less than a day) causing 10-fold changes in volume concentration and threefold changes in abundance. A variety of explanations may be offered for these observations, the simplest of which is the advection of water masses with different particle characteristics. Variation in temperature, current speed or direction recorded 80 m above the camera were not correlated with changes in marine snow concentration, although there were substantial long-term changes in current direction and speed. Furthermore if the observed particles sink at rates of $\sim 100 \text{ m d}^{-1}$ (refs 7, 8), they would have been produced in the upper mixed layer within 100 km of the mooring even at the highest current speeds recorded. The temporal changes can thus be considered as representative of this particular oceanographic regime.

Knowledge of the concentration of marine snow in the world's oceans is sparse and is based on a small number of observations^{1–6}. Most of these observations were made in the Pacific and generally report only the abundance, a parameter which is very sensitive to the lower size limit of the recording system. Nevertheless the absolute values of abundance and volume concentration found here in the summer and autumn at 270 m are very similar to two recent records from the western Atlantic⁵, implying agreement between different techniques in similar environments. Previous reports may well have missed pulses of marine snow because they lacked temporal coverage, and conclusions drawn from them should be treated with caution.

We believe that the observed variability reflects changes over time in the hydrographic and biological processes in the upper water column that are responsible for the production of marine snow. The Northeast Atlantic is an area of intense seasonality, and the seasonal marine snow signal clearly reflects trends in

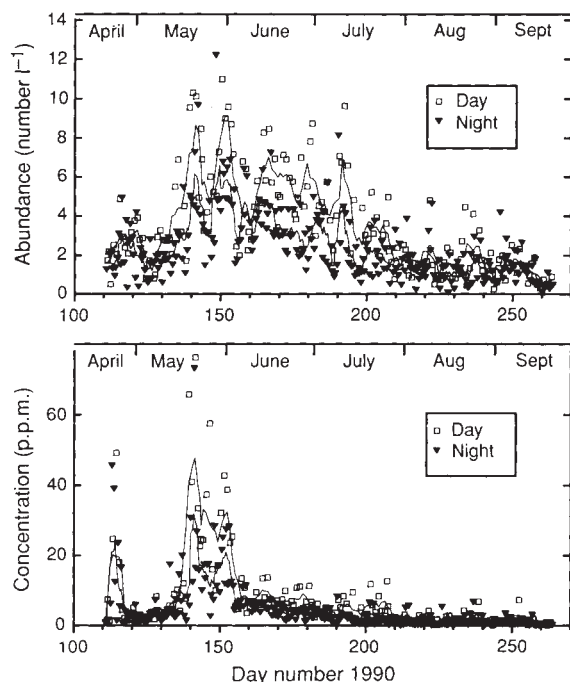


FIG. 1 Abundance and volume concentration of marine snow particles in the size range 0.6–9.9 mm equivalent spherical diameter during the deployment period. 'Day' (0500–1400 h GMT) and 'night' data are treated separately to reveal the pronounced nocturnal depression. Five-point running means have been applied to these data to give the solid lines. The Marine Snow camera system uses 35-mm Tmax film to photograph 20 litres of water using orthogonal illumination collimated by a bank of 30×30 cm Fresnel lenses, a system similar to that used previously¹⁸. The negatives are viewed with a charge-coupled device (CCD) video camera with 500-line resolution, and the particles individually measured in two dimensions (maximum and minimum) using a Kontron Vidas image analyser. All particles are assumed to lie along the mid-plane of the 30-cm-thick light slab and hence 80 cm from the camera lens. Their volumes are calculated assuming a circular cross-section, and the data binned into six logarithmic size categories of equivalent spherical diameter (ESD). Particles of ESD 0.6–1.6 mm were recorded using a subsample of the photographic frame representing an *in situ* volume of 4 litres whereas particles of ESD 1.6–9.9 mm were determined using the entire 20-litre volume. A sine curve was fitted to the volume concentration data by least squares. This removed most of the energy from the peak in the spectrum at 1 cycle per day (Fig. 3), showing that the fit was appropriate. The phase (time of day) of the maximum was used to make the distinction between 'day' and 'night'.

primary production in the surface layer. At a nearby site, this increased from ~ 0.5 g carbon $m^{-2} d^{-1}$ to a maximum value of 2 g C $m^{-2} d^{-1}$ in the second half of May, subsequently decreasing to 0.6 g C $m^{-2} d^{-1}$ in the first half of June (G. Savidge, personal communication). The marine snow peak at the start of the deployment implies that a productivity pulse had just occurred. A close coupling is thus indicated between the upper 50 m, which supports almost all of the primary production, and the mesopelagic zone 200 m below in which many of the particles lost from the upper mixed layer are remineralized or repackaged. Such close coupling has already been demonstrated from traps which collect sedimenting material^{9–11}, although the temporal resolution of these is rarely less than 1 week, and from time-lapse photography of the sea floor⁷. The link with primary production has been less easy to establish directly, partly because of the inherent problems of measuring annual variations in production with sufficient time resolution. Data from the Sargasso Sea give some indication of this link, although the relationship is not always clear¹⁹. Although sediment traps provide a measure of material flux, they give no information on the way in which the material is transported to depth. Furthermore, in the upper 1,000 m of the water column, the collected samples are often

contaminated with animals that have swum into the collecting cups. The present observations support the view that seasonal variations in material flux mainly depend on the larger sizes of marine snow, the smaller size categories showing little seasonal variation.

Even more surprising than the observed rapid and large seasonal changes in marine snow concentration are fluctuations within a day. Figure 3 depicts two periods of time when such diel variability was particularly pronounced. When marine snow concentrations were low in the late summer, the diel periodicity was less apparent, but for the entire data set and for all particle size classes there is a substantial peak in the power spectrum with a period of 24 hours. The diel periodicity is even apparent in the very small particles (<0.6 mm) for which no accurate dimensional data can be obtained. Current speed and direction vary with a typical lunar period (1.9 cycles per day) and are clearly not relevant to these observations. The most significant signal with a diel period at these depths is that of vertical migration of the plankton and nekton. Large numbers of these organisms traverse the upper few hundred metres daily, moving into the upper wind-mixed layer at night from depths as great as 1,000 m in the case of the larger nekton. Three explanations should be considered.

(1) Particles recorded as marine snow might in fact be the bodies of the migrating plankton and nekton. We do not consider this likely. Close examination of the 35-mm film revealed a negligible number of animals that would have been recorded as marine snow by the image analyser. Over the past 50 years, the volume concentration of these organisms (or biovolume) has been measured throughout the North Atlantic using nets of various types and during different times of the year and day. Except in productive inshore waters, the average over the top few hundred metres does not exceed ~ 150 ml per $1,000$ m³ (0.15 p.p.m.) with highest values at the surface and diel variability of less than half of this value^{12–14}. Nets do not efficiently collect fragile or fast-swimming animals, but it is unlikely that

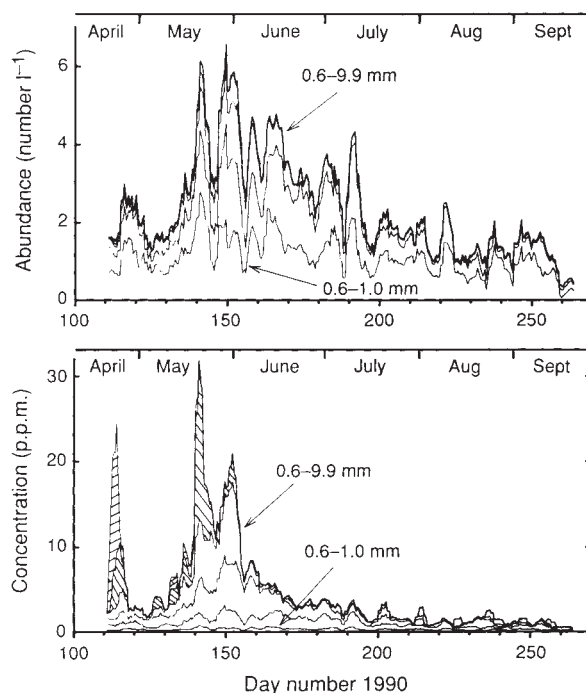


FIG. 2 Variation in cumulative abundance (upper) and volume concentration (lower) during the 'night' in each of the six size categories greater than 0.6 mm (<1.0 mm, <1.6 , <2.5 , <3.9 , <6.3 and <9.9 ; five-point running means). The cross shading identifies the contribution due to the two largest size categories (3.9–6.6 mm and 6.6–9.9 mm) which are particularly important at the beginning of major pulses of material.

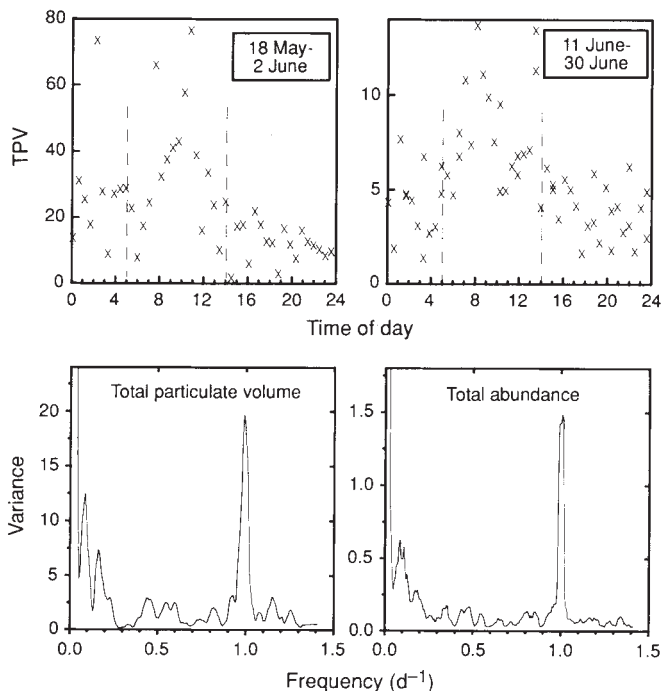


FIG. 3 Upper: accumulated data from two selected periods of time to demonstrate the diel variation in total particulate volume TPV (p.p.m.). Note the change of scale. The vertical dotted lines indicate the boundaries between 'day' and 'night'. Lower: power spectra for variation in total particulate-volume and total abundance over the entire deployment period. The power spectrum shows the amount of variance at each frequency. The peak at 1 cycle per day clearly shows the diel signal. The lower 95% confidence interval on this peak is at 0.48 times the height of the peak. This is substantially above the background spectrum, showing that the diel signal is significant.

the method underestimates volume concentration by more than 50% (0.08 p.p.m.). The volume concentration of marine snow varies by up to 30 p.p.m., particularly during periods dominated by large particles (Fig. 3), and even in the autumn period, the maximum plankton concentration likely is only 10% of the marine snow concentration (August and September mean: 1.4 p.p.m.). Finally, it is possible that plankton are attracted to the structure leading to higher concentrations in the field of view than elsewhere. But ship-based vertical profiles of marine snow near to the mooring site in May 1990 give very similar concentrations to those recorded on the mooring¹⁵. As the animals would not be able to swim as fast as the camera was moving, this suggestion may be discounted. For all these reasons, mis-identification of animals as marine snow is unlikely.

(2) A synchronous pulse of marine snow might leave the upper water column daily. Marine snow is likely to be produced in the upper water column as a result of the local biological and physical processes, both of which exhibit strong diel periodicity. The particles so produced are, however, likely to have a wide range of sinking rates⁸, from <10 to >200 m d^{-1} . If marine snow production is concentrated in the upper water column (<60 m depth), such a pulse of export would no longer be coherent by the time the material reached 270 m, the depth of the camera. It is conceivable that the marine snow pool at this depth is supplied from a small subset of the upper-ocean marine snow that has a very high and uniform sinking rate, and that this breaks up at the camera depth.

(3) The migrating plankton might produce and consume marine snow particles at a very high rate. There is increasing evidence that mid-water organisms feed on, break up and produce a variety of types of inanimate mid-water particle such as marine snow^{1,15-17}. The rates at which these processes occur and the role of different faunal groups are largely unknown. Fragile organisms that are poorly recovered in nets are certainly prin-

cipal sources of marine snow in some regions, and a number of the marine snow images may well be discarded appendicularian houses. It may be deduced that the diel variability of marine snow is intimately related to the activity of the migrating plankton and nekton. This does pose a difficult problem, in that the zooplankton and nekton must be able to cause changes to the particle pool of a much greater size than their biovolume. This has not so far been reported, but the possibility should be examined.

The physical nature of the particle pool clearly changes rapidly, sometimes over periods of only a few hours. Recent work suggests that released or surface-bound bacterial enzymes may cause rapid loss of organic matter from marine snow, giving turnover times of 0.2–2.1 days². Our data indicate that the free-living migrating biota may be responsible for short-term changes to the marine snow pool and hence also for the longer-term changes. We believe that they will substantially influence the seasonal cycle of marine snow, thus affecting the rates and ways in which material is transformed and deposited in the ocean. The relative importance of the attached and free-living fauna in marine snow dynamics remains an open question. □

Received 28 October 1992; accepted 17 March 1993.

1. Alldredge, A. L. & Silver, M. W. *Prog. Oceanogr.* **20**, 41–82 (1988).
2. Smith, D. C., Simon, M., Alldredge, A. L. & Azam, F. *Nature* **359**, 139–142 (1992).
3. Gotschalk, C. C. & Alldredge, A. L. *Mar. Biol.* **103**, 119–129 (1989).
4. Asper, V. L., Honjo, S. & Orsi, T. H. *Deep Sea Res.* **39**, 939–952 (1992).
5. Walsh, I. D. & Gardner, W. D. *Deep Sea Res.* (in the press).
6. Gardner, W. D. & Walsh, I. D. *Deep Sea Res.* **37**, 401–412 (1990).
7. Lampitt, R. S. *Deep Sea Res.* **32**, 885–897 (1985).
8. Alldredge, A. L. & Gotschalk, C. *Limnol. Oceanogr.* **33**, 339–351 (1988).
9. Deuser, W. G., Ross, E. H. & Anderson, R. F. *Deep Sea Res.* **28**, 495–505 (1981).
10. Deuser, W. G. et al. *Deep Sea Res.* **37**, 1331–1343 (1990).
11. Honjo, S. & Manganini, S. J. *Deep Sea Res.* **40**, 587–607 (1993).
12. Be, A. W. H., Forns, J. M. & Roels, O. A. in *Fertility of the sea* (ed. Costlow, J. D.) 17–50 (Gordon and Breach, New York, 1971).
13. Angel, M. V. & Baker, A. de C. *Biol. Oceanogr.* **2**, 1–30 (1982).
14. Angel, M. V. *Prog. Oceanogr.* **22**, 1–46 (1989).
15. Hamner, W. M., Madin, L. P., Alldredge, A. L., Gilmer, R. W. & Hamner, P. P. *Limnol. Oceanogr.* **20**, 907–917 (1975).
16. Lampitt, R. S., Noji, T. & von Bodungen, B. *Mar. Biol.* **104**, 15–23 (1990).
17. Lampitt, R. S., Wishner, K. F., Turley, C. M. & Angel, M. V. *Mar. Biol.* (in the press).
18. Honko, S., Doherty, K. W., Agrawal, Y. C. & Asper, V. L. *Deep Sea Res.* **31**, 67–76 (1984).
19. Asper, V. L., Deuser, W. G., Knauer, G. A. & Lohrenz, S. E. *Nature* **357**, 670–672 (1992).

ACKNOWLEDGEMENTS. We thank M. V. Angel, M. J. R. Fasham and M. W. Silver for comments on the manuscript, T. D. Jickells and P. P. Newton for logistic help, K. M. Goy, N. J. Griffin and C. Washington for technical assistance, and the master and crew of RRS *Charles Darwin*. This work was partially supported by the NERC UK JGOFS program Biogeochemical Ocean Flux Study.

Tracing trace elements from sediment input to volcanic output at subduction zones

Terry Plank* & Charles H. Langmuir

Lamont-Doherty Earth Observatory and Department of Geological Sciences at Columbia University, Palisades, New York 10964, USA

At ocean trenches, sea-floor sediments may either be scraped off the subducting plate, or accompany it into the mantle. Some of the subducted sediment may then be recycled to the arc crust by magmatism¹; the rest may be recycled into the mantle, and contribute to mantle heterogeneity². Strong evidence for sediment contributions to arc volcanism has come from isotope tracers, such as ²⁰⁷Pb and ¹⁰Be (refs 3–5), but a global mass balance requires consideration of element fluxes⁶. Here we report the sedimentary fluxes into eight trenches around the globe of trace elements that are enriched in arc volcanics (Ba, Sr, K, Rb, Cs, La, Th and U)⁷.

* Present address: Department of Geological Sciences, Snee Hall, Cornell University, Ithaca, New York 14853, USA.

TABLE 1 Sediment fluxes into trenches and arc basalt compositions

Arc	Guatemala	Mexico	Java	Tonga	East Aleutians	North Antilles	Vanuatu	Marianas
Convergence rate (cm yr ⁻¹)	6.4	6.5	8.1	12.2	5	3.7	8.6	7.7
Sediment thickness (m)	400	200	300	100	500	250	650	400
+/- (m)	100	50	100	50	100	50	100	100
Density (g cm ⁻³)	1.62	1.37	1.65	1.29	1.71	1.66	1.6	1.93
Water (wt%)	49	59	40	64	38	37	38	23
<i>Bulk compositions of sediment sections</i>								
K ₂ O (wt%)	0.67	1.23	2.50	2.09	2.22	1.81	2.29	0.95
Ba (p.p.m.)	3,250	2,175	1,068	2,105	1,582	204		306
Sr (p.p.m.)	1,237	234	218	261	234	111	293	76.4
Rb (p.p.m.)	21.8	49.4	88	55.4	59.7	84	20	26.5
Cs (p.p.m.)	1.31	3.2	4.9	3.18	3.47	4.89		1.5
La (p.p.m.)	16.83	42.4	39.08	110.2	19.36	35.54		13.5
U (p.p.m.)			1.47	1.46	2.58	1.36		0.45
Th (p.p.m.)	1.42	8.6	9.8	9.57	5.92	12.1		2.23
<i>Sediment fluxes (g yr⁻¹ per cm arc length)</i>								
K flux	1,176	746	4,960	983	4,885	1,454	10,542	3,610
Ba flux	687	159	255	119	419	20		140
Sr flux	261.6	17.1	52.1	14.8	62.0	10.7	162.5	35.0
Rb flux	4.61	3.61	21.03	3.14	15.82	8.13	11.09	12.13
Cs flux	0.28	0.23	1.17	0.18	0.92	0.47		0.69
La flux	3.56	3.10	9.34	6.24	5.13	3.44		6.18
U flux			0.35	0.08	0.68	0.13		0.21
Th flux	0.30	0.63	2.34	0.54	1.57	1.17		1.02
<i>Average basaltic compositions</i>								
K 6.0	0.71	1.02	1.01	0.35	0.77	0.53	0.95	0.61
Ba 6.0	397	302	269	122	298	108	206	163
Sr 6.0	539	526	340	195	460	360	584	310
Rb 6.0	11.9	13.2	21.4	4.2	12.6	5.4	16.9	9.5
Cs 6.0		0.56	0.94	0.20	0.51	0.16		0.32
La 6.0	6.7	10.8	11.6	1.2	8.2	5.4	7.5	4.9
U 6.0		0.42	0.64	0.12	0.63	0.41	0.41	0.24
Th 6.0	0.99	0.91	2.95	0.13	1.56	1.10	1.00	0.56
Na 6.0	2.90	3.87	2.98	1.57	2.87	2.76	2.53	2.39
Number of volcanoes	11	5	6	5	6	4	4	8

Convergence rates (normal to the arc) are from ref. 23 based on the model in ref. 30. Sediment thickness, density and percentage water are based on ODP and DSDP initial reports (sites given in Fig. 1 and ref. 20). Bulk sediment compositions are calculated²⁰ from sediment analyses in refs 5, 17, 31–37. Sediment fluxes calculated as the product of convergence rate, thickness, density (100 – % water) and element concentration, where % water is the weight loss at 110 °C and element concentrations are reported relative to weight at 110 °C. Arc compositions at 6% MgO are calculated for basalts >5% MgO (data sources in ref. 20). Values are averages of individual volcanoes; number of volcanoes per arc (*n*) as indicated; number of analyses varies for each volcano and element²⁰.

We show that the volcanic outputs clearly reflect the sediment inputs, once the effects of melting are taken into account⁸. Where the sediment flux into a trench is high for a particular element, the associated volcanics are enriched in this same element. Thus, some of the geochemical characteristics of arc volcanics can be traced back to the sediments at the trench. A mass balance of the inputs and outputs will ultimately provide estimates for how much sediment is recycled to the arc, and how much to the deeper mantle.

Calculating the flux of subducted sediment into a given trench is straightforward. Convergence rates can be estimated from plate motion models, and sediment masses can be estimated from the thickness and density of drilled sedimentary sections near trenches. An additional consideration, however, is the extent of sediment accretion in the fore-arc, for accreted sediments are not being subducted. For several trenches (for example, Aleutians, Antilles, Mexico), a décollement separates sediments that are being accreted from those that are being subducted^{9–11}. For other trenches (for example, Tonga, Marianas, Guatemala), the fore-arc lacks oceanic sediment, and the entire sedimentary section is apparently being subducted^{12–14}. The sedimentary columns (Fig. 1) and mass fluxes (Table 1) presented here are only for those sediments which are demonstrably being subducted.

To convert mass fluxes to chemical fluxes, the bulk chemical composition of sedimentary columns must be estimated. The geochemical variability within sedimentary columns may be

extreme: sediment cores that are hundreds of metres in length often preserve variability on the scale of centimetres. Our approach takes advantage of the close correspondence between geochemical and lithological variations in marine sediments^{15,16}, and the abundant lithological data that are published for deep-sea drill cores. For example, biogenic phases (opal and calcium carbonate) are barren of most trace elements (such as alkalis, Th, rare earth elements), and thus are important diluents. Biogenic dilution is an important aspect of the chemical flux because much of the subducted mass consists of biogenic clays and oozes (Fig. 1). Dilution trends can be characterized with few chemical analyses, and the bulk percentage of diluting phases can be estimated from published lithological data. From this information alone, the bulk composition for some sedimentary columns can be calculated accurately¹⁷. Other simplifying systematics exist for elements that have nearly constant, continental ratios¹⁸ (such as Th/Al), and elements that are linked to other marine phases (for example rare earth elements to fish teeth)¹⁹. Further details and tests of our methods for different reference drill sites are presented elsewhere, along with the chemical analyses of over 250 sediment samples (refs 17, 20; and T.P. and C.H.L., manuscript in preparation).

The calculated fluxes into oceanic trenches vary by more than an order of magnitude for many elements (Table 1). Do these large variations in the sediment inputs to subduction zones affect the volcanic outputs? To answer this question, we compiled

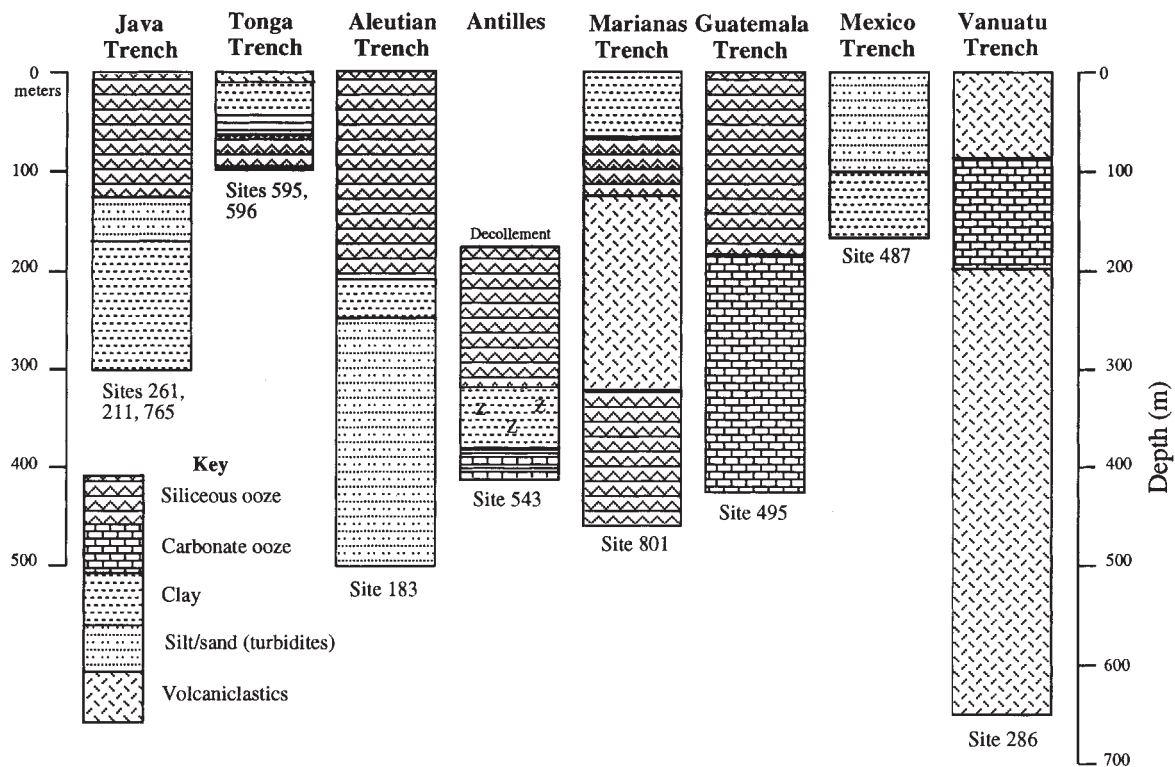


FIG. 1 Average sedimentary columns being subducted at eight trenches. Columns are estimated from deep-sea drilling and seismic reflection profiles (details and data sources in ref. 20; and T.P. and C.H.L., manuscript in

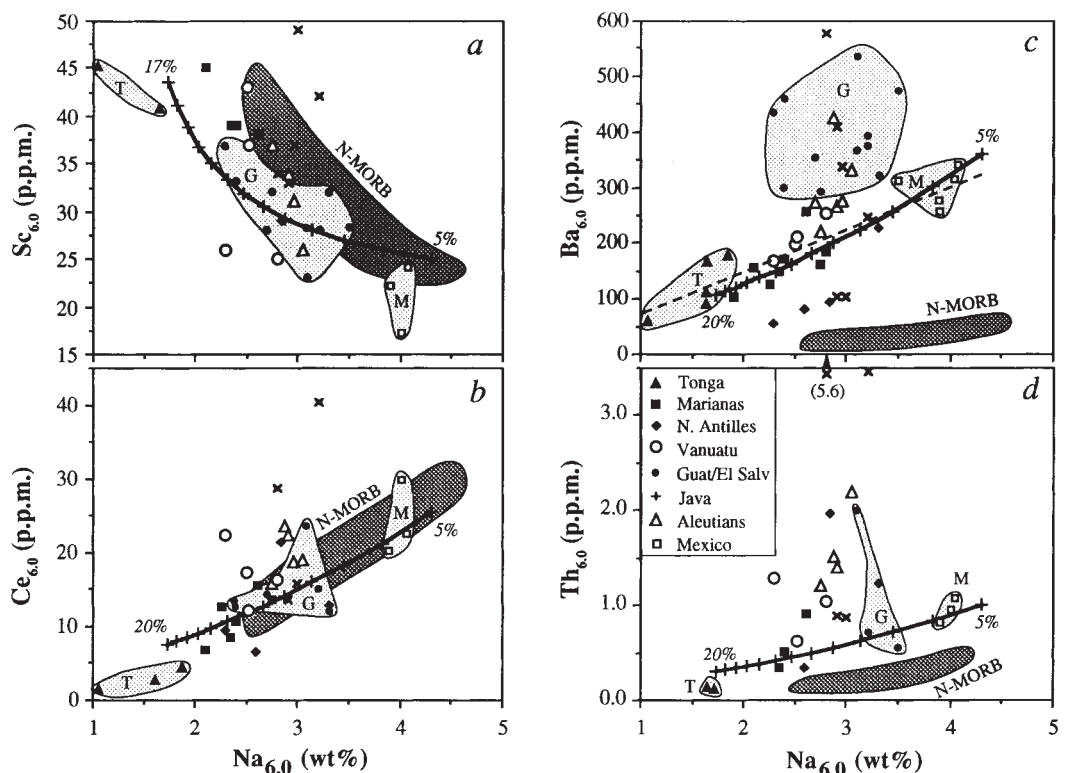
chemical data for arc volcanics from the same eight subduction zones for which we were able to calculate sediment input fluxes.

Arc volcanics have a complex history, involving crustal differentiation and mantle melting, and these effects need to be

preparation). Lithologies are based on reference DSDP or ODP drill sites (as indicated). Patterns as in key and as follows: solid horizontal rule (metalliciferous sediment); filled siliceous pattern (chert); and Z (zeolites).

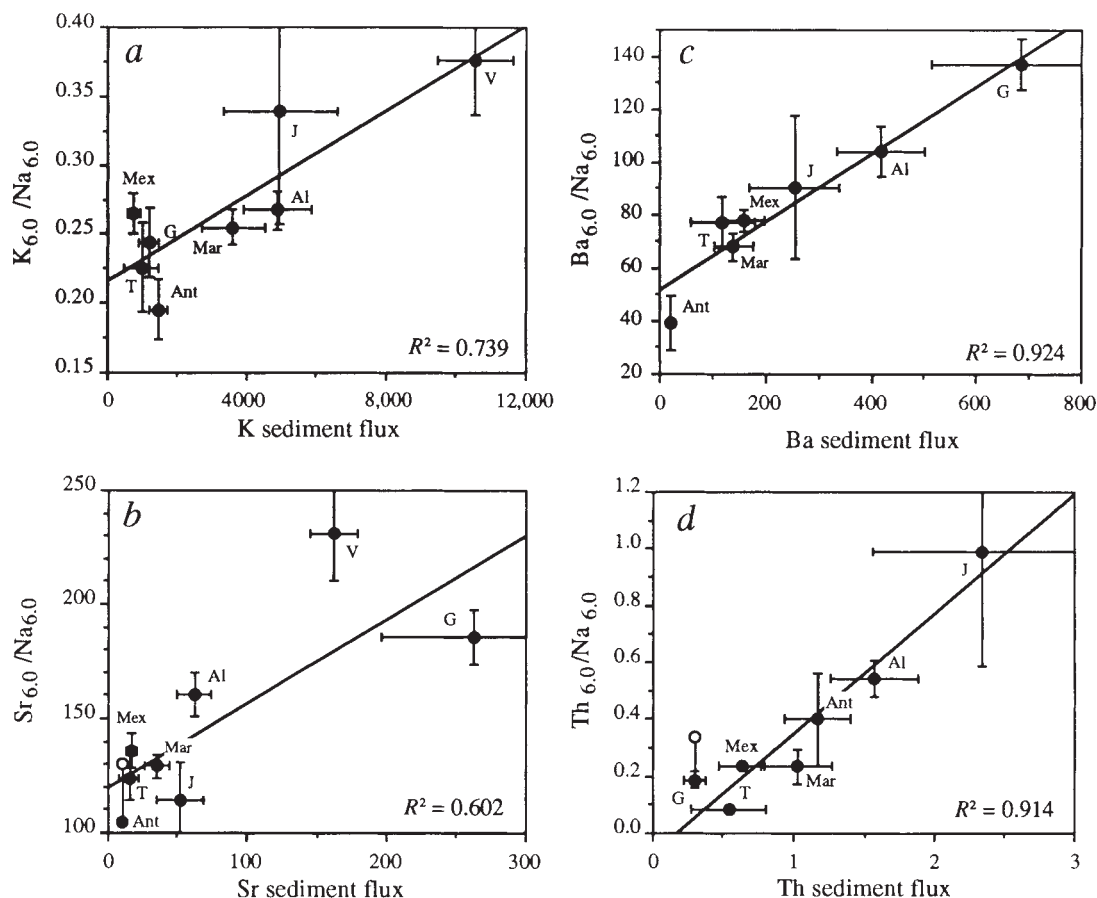
taken into account if we are to see through them to the influence of subducted sediment in the source region. The effects of crustal processes (such as polybaric crystallization and crustal assimilation) can be minimized by comparing volcanic compositions at

FIG. 2 Arc basalt and normal mid-ocean-ridge basalt (N-MORB) compositions at 6 wt% MgO. Arc points are for individual volcanoes from the same eight arcs for which sediment fluxes have been calculated (Table 1; see ref. 20 for data sources). Volcanoes from Tonga, Mexico and Guatemala/El Salvador arcs are highlighted. N-MORB field is from refs 22 and 37 where compositions at 8% MgO are corrected to 6% MgO by means of the following expressions: $\Delta\text{Na}/\Delta\text{Mg} = (-0.3)$, $\Delta\text{Sc}/\Delta\text{Mg} = 3.33$, $\Delta\log(\text{Ce}, \text{Ba}, \text{Th})/\Delta\text{Mg} = (-0.11)$. Melting curves (percentage melting indicated) are for nonmodal melting for Sc ($K_{\text{d}_{\text{cpx}}} = 4$, $P = 0.6$, $D_0 = 0.6$, $C_0 = 16$ p.p.m. where $K_{\text{d}_{\text{cpx}}}$ is the clinopyroxene/liquid partition coefficient, P is the partition coefficient for the solid entering the liquid, D_0 is the initial bulk partition coefficient, and C_0 is the initial concentration) and batch melting for Na ($D = 0.03$, $C_0 = 0.3$), Ce ($D = 0$, $C_0 = 0.96$ p.p.m.), Ba ($D = 0$, $C_0 = 13.5$ and 2.25 p.p.m.) and Th ($D = 0$, $C_0 = 0.0375$). Melting curves are corrected to 6% MgO assuming $F = 0.75$, $D = 0$ for Ce, Ba, Th; $D = 0.6$



for Na; and $D = 1.8$ for Sc. Dashed line in c shows a reference Ba/Na ratio of ~ 75 .

FIG. 3 Correlations between sediment fluxes into trenches and enrichments in arc basalts. Sediment fluxes (in g yr^{-1} per cm arc length) are from Table 1. $\text{Na}_{6,0}$ -normalized ratios are averages of basalts from each arc (Table 1 and Fig. 2), and reflect the relative enrichment in the source region of each arc. Horizontal error bars represent uncertainties in sediment thickness (as in Table 1); vertical error bars are $\pm$ one standard deviation of the mean of volcanoes within each arc; R^2 values are for linear regressions. In *b* and *d*, open circles are averages with and solid circles are averages without Redonda (Antilles) and Toliman (Guatemala) volcanoes, anomalous volcanoes for these two arcs.



a high MgO concentration. In general, the lower the MgO, the more a magma has crystallized, and thus the more heat it has released with which to melt and assimilate wallrock. For this reason, only the most primitive arc basalts (>5 wt% MgO) are considered here, and average compositions for each volcano are calculated at a reference value of 6 wt% MgO (ref. 8). For example, $\text{Ba}_{6,0}$ is the average Ba concentration at 6% MgO for basaltic samples from a single volcano. This Ba index is not equal to the composition of the primitive magma itself, but reflects a common reference point with which to compare the overall level of Ba enrichment or depletion of the magma that supplies each volcano.

The effects of partial melting in the mantle on the compositions of arc basalts must also be removed. For example, Na, Ca, Ce and Sc variations in arc basalts are well explained by different extents of mantle melting⁸ (Fig. 2*a* and *b*). The importance of melting on the global variability of arc volcanics is also supported by the similarity in both the absolute values and the trends of these elements to those in mid-ocean-ridge basalts^{21,22} (Fig. 2*a* and *b*), for which crustal processes are much less invasive and mantle melting is better understood²². Melting effects account for roughly a factor of 3 variation in the abundances of incompatible elements in arc volcanics⁸. To remove such melting effects from the geochemical signal, we normalize to $\text{Na}_{6,0}$, which appears to vary inversely with the extent of melting^{8,21}. Ba and Th (and the other elements in Table 1) are enriched in arc basalts relative to mid-ocean-ridge basalts and vary too much to be explained by melting alone (Fig. 2*c* and *d*). The $\text{Ba}_{6,0}/\text{Na}_{6,0}$ ratios of arc basalts are thus taken to reflect the Ba enrichment in the source, which may derive from the subducted plate.

Variations in the Na-normalized ratios of arc basalts correlate well with variations in the sediment fluxes (Fig. 3*a-d*), suggesting that arc outputs reflect sediment inputs. Indeed, some of the perplexing characteristics of arc volcanics can be traced

back to the subducting sediments. For example, Ba and Th are both highly incompatible elements and would be expected to be enriched or depleted together as a result of differences in the extent of melting or mantle heterogeneity. Guatemalan volcanics, however, have high Ba and low Th. These characteristics seem to originate in the biogenic oozes that are being subducted at the Guatemala trench, which are rich in Ba because of barite precipitation onto biogenic material, but poor in Th and other detrital elements (K, Rb) because of dilution by biogenic material.

Sediment subduction explains many other geochemical characteristics of arcs that would not have been predicted otherwise. For example, the Na-normalized ratios of Tongan and Mexican basalts are similar (Figs 2 and 3), suggesting similar sources for the two arcs. This similarity is surprising given the very different physical settings for the two arcs: the Tonga arc is built on thin oceanic crust where ancient crust subducts and a young back-arc basin has developed, whereas the Mexican volcanic arc is built on thick continental crust over a shallowly dipping, young plate²³. Other models that have been proposed to explain trace element variability among arcs, such as previous depletion in the mantle wedge^{24,25}, sub-continental lithosphere^{26,27} or crustal assimilation^{28,29}, would fail to explain the geochemical similarity in Tongan and Mexican sources. The sediment subduction model succeeds because subducted fluxes are similar for both arcs, primarily because of short sedimentary columns (Fig. 1).

The correlations between sediment inputs and arc enrichments emerge only after careful calculation of sediment fluxes and after correction of the volcanic data for melting effects. Thus the correlations provide additional evidence for the importance to arc volcanism of different extents of melting in the mantle⁸. The correlations also show that elements themselves can be traced through subduction zones. This result enables us to move from generic models for sediment involvement based on isotope

and trace element ratios to more quantifiable models based on element fluxes. As we show elsewhere (ref. 20, and T.P. and C.H.L., manuscript in preparation), this mass balance indicates that ~20% of the element budget in subducted sediment is recycled to the arc. Although such recycling plays an important role in arc volcanism, a larger fraction of subducted sediment may continue to descend with the plate into the deeper mantle. □

Received 29 December 1992; accepted 4 March 1993.

1. Armstrong, R. L. *Earth planet. Sci. Lett.* **12**, 137–142 (1971).
2. White, W. M. *Geology* **13**, 115–118 (1985).
3. Kay, R. W., Sun, S.-S. & Hu, C.-N. *Geochim. cosmochim. Acta* **42**, 263–273 (1978).
4. Tera, F., Brown, L., Morris, J., Sacks, I. S., Klein, J. & Middleton, R. *Geochim. cosmochim. Acta* **50**, 535–550 (1986).
5. White, W. M., Dupre, B. & Vidal, P. *Geochim. cosmochim. Acta* **49**, 1875–1886 (1985).
6. Karig, D. E. & Kay, R. W. *Phil. Trans. R. Soc. Lond. A* **301**, 233–251 (1981).
7. Gill, J. B. *Orogenic Andesites and Plate Tectonics* (Springer, New York, 1981).
8. Plank, T. & Langmuir, C. H. *Earth planet. Sci. Lett.* **90**, 349–370 (1988).
9. von Huene, R. & Scholl, D. W. *Rev. Geophys.* **29**, 279–316 (1991).
10. Westbrook, G. K., Ladd, J. W., Buhl, P., Bangs, N. & Tiley, G. J. *Geology* **16**, 631–635 (1988).
11. Moore, G. F. & Shipley, T. H. *J. geophys. Res.* **93**, 8911–8927 (1988).
12. Bloomer, S. H. & Fisher, R. L. *J. Geol.* **95**, 469–495 (1986).
13. Hussong, D. M. & Uyeda, S. *Init. Rep. DSDP* **60**, 909–929 (1982).
14. von Huene, R. & Aubouin, J. *Init. Rep. DSDP* **67**, 775–796 (1982).
15. Heath, G. R. & Dymond, J. *Geol. Soc. Am. Bull.* **88**, 723–733 (1977).

16. Kyte, F. T., Leinen, M., Heath, G. R. & Zhou, L. *Geochim. cosmochim. Acta* (in the press).
17. Plank, T. & Ludden, J. N. *Proc. ODP Sci. Res.* **123**, 167–189 (1992).
18. Taylor, S. R. & McLennan, S. M. *The Continental Crust: its Composition and Evolution* (Blackwell, Oxford, 1985).
19. Toyoda, K., Nakamura, Y. & Masuda, A. *Geochim. cosmochim. Acta* **54**, 1093–1103 (1990).
20. Plank, T. thesis, Columbia Univ. (1993).
21. Klein, E. M. & Langmuir, C. H. *Geophys. Res.* **92**, 8089–8115 (1987).
22. Langmuir, C. H., Klein, E. M. & Plank, T. *AGU Monogr.* (in the press).
23. Jarrard, R. D. *Rev. Geophys.* **24**, 217–284 (1986).
24. Ewart, A. & Hawkesworth, C. J. *J. Petrol.* **28**, 495–530 (1987).
25. McCulloch, M. T. & Gamble, J. A. *Earth planet. Sci. Lett.* **102**, 358–374 (1991).
26. Pearce, J. A. *Continental Basalts and Mantle Xenoliths* (eds Hawkesworth, C. J. & Norry, M. J.) 230–249 (Shiva, Nantwich, 1983).
27. Varne, R. & Foden, J. D. *The Origin of Arcs* (ed. Wezel, F.-C.) 159–189 (Elsevier, Amsterdam, 1986).
28. Leeman, W. P. J. *Volc. geotherm. Res.* **18**, 561–588 (1983).
29. Hildreth, W. & Moorbath, S. *Contrib. Mineral. Petrol.* **98**, 455–489 (1988).
30. Minster, J. B. & Jordan, T. H. *J. geophys. Res.* **83**, 5331–5354 (1978).
31. Kay, R. W. & Kay, S. M. *Geochim. cosmochim. Acta* **52**, 1351–1359 (1988).
32. Carr, M. J., Feigenson, M. D. & Bennett, E. A. *Contrib. Mineral. Petrol.* **105**, 369–380 (1990).
33. Davidson, J. P. *Geochim. cosmochim. Acta* **51**, 2185–2198 (1987).
34. Karl, S. M., Wandless, G. A. & Karpoff, A. M. *Proc. ODP Sci. Results* **129**, 31–80 (1992).
35. Leggett, J. K. *Init. Rep. DSDP* **66**, 683–686 (1982).
36. Briquieu, L. & Lancelot, J. R. *Geochim. cosmochim. Acta* **47**, 191–200 (1983).
37. Zhou, L. & Kyte, F. T. *Paleoceanography* **7**, 441–465 (1992).
38. Plank, T. & Langmuir, C. H. *J. geophys. Res.* **97**, 19749–19770 (1992).

ACKNOWLEDGEMENTS. We thank M. J. Carr for unpublished data; L. Zhou, F. Kyte and S. Karl for preprints of their manuscripts; W. M. White, R. W. Kay and S. M. Kay for splits of their samples; J. Morris, J. Rubenstein, A. Zindler, D. Walker and T. Elliott for useful discussions; and A. Hofmann, M. J. Carr and J. Rubenstein for reviews. This work was supported by a JOI/USSAC Graduate Fellowship and the U.S. NSF.

Learning to recognize nestlings is maladaptive for cuckoo *Cuculus canorus* hosts

Arnon Lotem*

Department of Zoology, Faculty of Life Sciences, Tel Aviv University, Ramat Aviv, Tel Aviv 69978, Israel

THE picture of a tiny passerine host feeding a huge cuckoo nestling challenges evolutionary biologists who explain animal behaviour as adaptive^{1–4}. Cuckoo eggs sometimes resemble the eggs of the host, but nestlings of the common cuckoo, *Cuculus canorus*, look very different from the young of the host. The inability of the host to discriminate against such divergent nestlings is especially puzzling as some cuckoo hosts show a finely tuned discrimination ability between eggs^{5–8}. Here I present a simple model to explain this paradox. The model shows that although learning to recognize eggs is adaptive, learning to recognize nestlings might not be. The mechanism of learned recognition, previously shown to maintain egg recognition, is unlikely to be adaptive for hosts like those of the common cuckoo, in which only the parasitic nestling remains in the nest. The reason that discrimination against parasite nestlings is not adaptive is that the cost of misimprinting (learning to recognize the parasite nestling as the parents' own) exceeds the benefit of correct learning. The model also explains why nestling discrimination is mostly found in host-parasite systems in which the parasite and the hosts' young are reared together¹.

Recent studies indicate that cuckoos and their hosts may have evolved several adaptations and counter adaptations to each other^{1,5,9,10}. Egg recognition is the common host defence against cuckoo parasitism, but some level of acceptance of cuckoo eggs by hosts is also common^{6,8,9}. Mimicry in cuckoo eggs and the cost of errors in egg recognition may prevent the parasitized host from always rejecting the cuckoo eggs⁷. But because cuckoo nestlings look different from the hosts' young, nestling discrimination should be easy. Soon after hatching, the cuckoo's nestling ejects all host eggs or nestlings from the nest. Although it is thus too late for the host to save its young, by deserting the

cuckoo nestling a host can save about 30 days of parental care which might be used for re-nesting. It is therefore surprising that hosts have not developed nestling recognition by the same mechanism used for recognizing eggs.

Several studies suggest that egg recognition in birds is learned by an imprinting-like process^{7,11–13}. Accordingly, hosts imprint on their own egg type during their first breeding attempt and later will reject any different egg type. Given this ability, a host should be able to learn the appearance of its nestlings when it first breeds, and use this knowledge to discriminate against parasitic nestlings in future breeding attempts. But, as I show below, learning to recognize nestlings is maladaptive for hosts like those of the common cuckoo, for which only the parasitic nestling survives in parasitized nests. Such hosts, if parasitized during their first breeding attempt, will face only a parasitic nestling during their learning period, and consequently will reject their own young in any future breeding attempt. In other

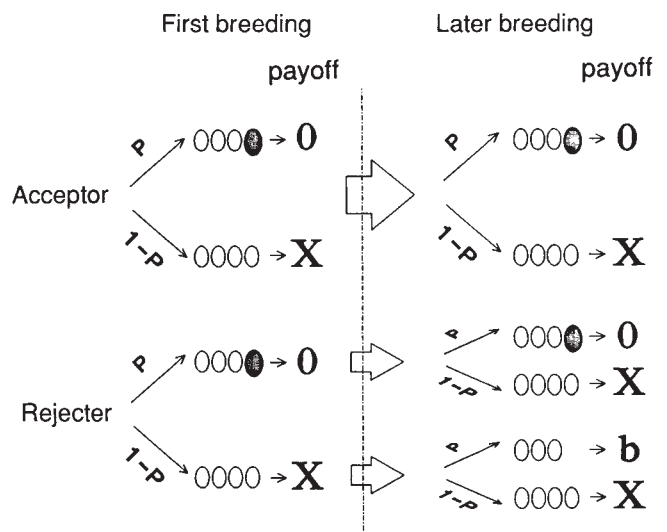


FIG. 1 The mechanism of learned egg recognition, for a typical host of a common cuckoo (see text for explanation). *P*, the host probability of being parasitized; *X*, the average reproductive success (number of fledglings) of a non-parasitized host; 0, zero: the reproductive success from a parasitized nest; *b*, the benefit of egg rejection, which is the average reproduction success of a parasitized host that rejects the cuckoo egg.

* Present address: Department of Zoology, University of British Columbia, 6270 University Boulevard, Vancouver, BC V6T 1Z4, Canada.

words, the learning mechanism may incur a cost, termed here the 'misimprinting cost', which is the result of learning to recognize the wrong type of egg or nestling as 'own'. The following model explores the effect of the misimprinting cost on the evolution of egg recognition (Fig. 1) and nestling recognition (Fig. 2) in a typical host of the common cuckoo (such as the great reed warbler, *Acrocephalus arundinaceus*).

Each arrow in the model is labelled with its probability of occurrence, and a payoff is listed at the end of each path. A rejection strategy is adaptive if its overall payoff is greater than that of an acceptance strategy. In either case (egg or nestling recognition), if an acceptor is parasitized by a common cuckoo (a probability of P), its reproductive success is usually zero^{14,15}, and if it is not parasitized (probability of $1-P$), its reproductive success is on average equal to x ($x > 0$). As an acceptor, by definition, does not learn to discriminate, its payoff does not change between its first breeding and the later breeding, and is constantly equal to $P0 + (1-P)x$, which is $(1-P)x$. In general, the payoff of an acceptor will be $n(1-P)x$, where n is the number of breeding cycles during a host lifetime.

The payoff for a rejecter is more complex and will be considered first for the case of egg recognition (Fig. 1). If a naive rejecter is parasitized during its first breeding, there is a risk not only that its payoff will be reduced to zero, but that it may also learn to recognize the cuckoo egg, as well as the other eggs in the clutch, as its own. Such an individual, if parasitized during later breeding attempts, might accept the cuckoo egg as its own, and consequently its payoff will be the same as that of an acceptor. However, if a rejecter is not parasitized in its first breeding and does learn to recognize its own type of eggs correctly, and if it is parasitized in later breeding, it will gain the benefit of rejection, designated as b . The payoff of a rejecter is therefore $P0 + P(n-1)(P0 + (1-P)x) + (1-P)x + (1-P)(n-1)(Pb + (1-P)x)$, which reduces to $P(n-1)((1-P)x) + (1-P)x + (1-P)(n-1)(Pb + (1-P)x)$. According to the model, and assuming that the parasitism rate is never 100% ($P < 1.0$), a rejecter will do better than an acceptor as long as $P(n-1)((1-P)x) + (1-P)x + (1-P)(n-1)(Pb + (1-P)x) > n(1-P)x$, which after solving, gives $Pb > 0$, or, put in words, rejection is favoured as long as the population is parasitized ($P > 0$) and that some benefits are gained by rejection ($b > 0$). Clearly, these are elementary conditions for the adaptiveness of

rejection *per se*, no matter what the rejection mechanism is. Hence, despite the cost of misimprinting, the proposed learning mechanism of egg recognition is adaptive.

The situation is different in the case of nestling recognition (Fig. 2). As in parasitized nests only the cuckoo nestling remains, a naive rejecter, if parasitized, may learn to recognize only the cuckoo nestling as its own. Such an individual will always accept cuckoo nestlings when parasitized, but because it was not exposed to its own young during the learning period, it will reject them in later breeding attempts. Hence, the cost of misimprinting in this case is the loss of all future reproductive success. Considering that cost, the payoff of a rejecter will be $P0 + P(n-1)(P0 + (1-P)0) + (1-P)x + (1-P)(n-1)(Pb + (1-P)x)$, which reduces to $P(n-1) + (1-P)x + (1-P)(n-1)(Pb + (1-P)x)$. Assuming that $0 < P < 1$ and $x > 0$, rejecters will do better than acceptors only if $P(n-1) + (1-P)x + (1-P)(n-1)(Pb + (1-P)x) > n(1-P)x$, which after solving, gives $b > x$. Put in words, rejecters will do better than acceptors only if the benefit of nestling rejection (b) is greater than the reproductive value of unparasitized nests (x). A situation like this probably never occurs; when rejection is achieved by deserting the cuckoo nestling, the benefit of rejection is likely to be lower than the value of an unparasitized nest because re-nesting is not always possible, and even when it occurs, later nests tend to have smaller clutches¹⁶ and to lead to lower reproductive success¹⁷. The benefit of rejection might be higher if the host can discriminate and eject the cuckoo nestling immediately after it hatches, thus saving its own young. But even if the host can save all its young ($b = x$) rejection is no better than acceptance. Moreover, because the cuckoo usually removes one host egg when laying its own^{1,15}, the host still remains with fewer nestlings than if it had not been parasitized ($b < x$). In conclusion, the proposed learning mechanism cannot be adaptive when the cuckoo nestling remains alone in the nest.

For simplicity, the model assumes equal probability of facing a cuckoo nestling in the first and in later breeding attempts. But in cases in which experienced breeders are more likely to reject cuckoo eggs⁷, they are inevitably less likely to face cuckoo nestlings. Consequently, the probability of facing a cuckoo nestling as a naive breeder, and thereby paying the cost of misimprinting, is greater than the probability of facing it as an experienced breeder and then gaining the benefit of correct learning. The likelihood of learned recognition being adaptive is thus even less than that predicted by the model.

The cost of misimprinting is also relevant to cases where the parasite and the host nestling are reared together. If host nestlings frequently fail to compete for food with the parasite and starve to death, the cost of misimprinting may outweigh the benefits of correct learning. On the other hand, if both host and parasite nestlings survive to fledging, learning to recognize nestlings should be adaptive. In agreement with this prediction, comparative evidence shows that nestling discrimination is mostly found in host-parasite systems in which the parasite is reared along with the host young¹ and causes little or no mortality among them^{9,18,19}. Although more comparative evidence of this kind can partially test the model, the mechanism of nestling discrimination needs to be studied directly. The learning hypothesis can be tested by experimentally replacing host nestlings with other nestling types during the host first breeding attempt. My model will be falsified if nestling discrimination is exhibited by hosts in which only the parasitic nestling remains in the nest, and if discrimination in this case is learned.

The model does not explain why hosts cannot recognize their nestlings innately. But it is doubtful whether the common forms of innate recognition could be sufficiently effective to reject a single cuckoo nestling. An innate response to key releasers is often stimulated by a broad range of stimuli that only slightly resemble the 'correct' one^{20,21}. Animals may prefer the stronger stimulus when faced with a simultaneous choice, but still respond to the weaker stimulus when it is the only one pro-

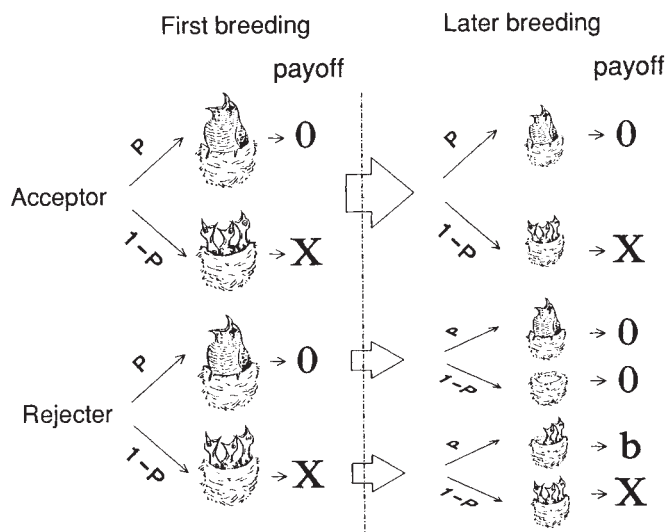


FIG. 2 Hypothetical mechanism of learned nestling recognition, for a typical host of a common cuckoo (see text for explanation). P , the host probability of being parasitized; X , the average reproductive success (number of fledglings) of a non-parasitized host; 0, zero: the reproductive success from a parasitized nest; b , the benefit of nestling rejection, which is the average reproduction success of a parasitized host that rejects the cuckoo nestling.

vided²¹. Such an innate preference may allow nestling discrimination where the parasite and the hosts' young are reared together, but not in common cuckoo hosts where the parasite remains alone. What is needed, but apparently does not exist, is a genetic programme that allows the hosts to ignore all other similar stimuli, and to desert only the single cuckoo nestling. The apparent lack of such a mechanism cannot serve as an explanation, but should stimulate further research into innate recognition and its limitations. □

Received 2 December 1992; accepted 16 February 1993.

1. Davies, N. B. & Brooks, M. de L. *Nim. Behav.* **36**, 262–284 (1988).
2. Davies, N. B. & Brooke, M. de L. *J. Anim. Ecol.* **58**, 225–236 (1989).
3. Dawkins, R. & Krebs, J. R. *Proc. R. Soc. Lond. B* **205**, 489–511 (1979).
4. Rothstein, S. I. *Am. Zool.* **22**, 547–560 (1982).
5. Brooke, M. de L. & Davies, N. B. *Nature* **335**, 630–632 (1988).
6. Davies, N. B. & Brooke, M. de L. *J. Anim. Ecol.* **58**, 207–224 (1989).
7. Lotem, A., Nakamura, H. & Zahavi, A. *Behav. Ecol.* **3**, 128–132 (1992).
8. Moksnes, A. *et al. Behaviour* **116**, 64–89 (1990).
9. Rothstein, S. I. *A. Rev. ecol. Syst.* **21**, 481–508 (1990).
10. Soler, M. & Møller, A. P. *Nature* **343**, 748–750 (1990).
11. Rothstein, S. I. *Auk* **91**, 796–807 (1974).
12. Rothstein, S. I. *Nim. Behav.* **26**, 671–677 (1978).
13. Victoria, J. K. *Ibis* **114**, 367–376 (1972).
14. Cramp, S. (ed.) *The Birds of the Western Palearctic* Vol. 4 (Oxford University Press, Oxford, 1985).
15. Wyllie, I. *The Cuckoo* (Batsford, London, 1981).
16. Klomp, H. *Ardea* **58**, 1–124 (1970).
17. Perrins, C. M. & Birkhead, T. R. *Avian Ecol.* (Blackie, Glasgow, 1983).
18. Payne, R. B. *A. Rev. ecol. Syst.* **8**, 1–28 (1977).
19. Payne, R. B. *Ecology* **58**, 500–513 (1977).
20. Lack, D. *The Life of the Robin* (Witherby, London, 1946).
21. Tinbergen, N. *The Study of Instinct* (Oxford University Press, London, 1951).

ACKNOWLEDGEMENTS. I thank N. Davies, A. Erez, O. Hasson, T. Hatfield, H. Nakamura, T. Redondo, S. Rothstein, D. Schluter, J. Smith, D. Ward and A. Zahavi for comments and suggestions. This work was supported by the United States-Israel Binational Science Foundation (BSF) and by an NSERC postdoctoral fellowship of the Canadian Government.

Male-driven evolution of DNA sequences

Lawrence C. Shimmin, Benny Hung-Junn Chang & Wen-Hsiung Li*

Center for Demographic and Population Genetics, University of Texas, PO Box 20334, Houston, Texas 77225, USA

It is commonly believed^{1,2} that the mutation rate is much higher in the human male germ line than in the female germ line because the number of germ-cell divisions per generation is much larger in males than in females. But direct estimation of mutation rates is difficult, relying mainly on sex-linked genetic diseases³, so the ratio (α_m) of male to female mutation rates is not clear. It has been noted⁴ that if α_m is very large, then the rate of synonymous substitution in X-linked genes should be only 2/3 of that in autosomal genes, and comparison of human and rodent genes supported this prediction⁴. As the number of X-linked genes used in the study was small and the X-linked and autosomal sequences were non-homologous, and given that the synonymous rate varies among genes⁵, we sequenced the last intron (~1 kb) of the Y-linked and X-linked zinc-finger-protein genes (ZFY and ZFX) in humans, orang-utans, baboons and squirrel monkeys. The ratio Y/X of the substitution rate in the Y-linked intron to that in the X-linked intron is ~2.3, which is close to that estimated from synonymous rates in the ZFY and ZFX genes^{6–8} and implies $\alpha_m \approx 6$. This estimate of α_m supports the view that the evolution of DNA sequences in higher primates is male-driven. It is, however, much lower than the previous estimate⁴ and therefore raises a number of issues.

Figure 1 shows a condensed alignment of the sequences of the last intron of the human, orang-utan, baboon and squirrel

monkey ZFX and ZFY genes; only the human X and Y sequences are presented. We shall restrict our analysis to the homologous parts of the X and Y introns, excluding the Alu sequence, which is present in ZFX but not in ZFY.

Table 1 shows the corrected per cent divergence between sequences. For each species pair, the Y sequences are more divergent than the X sequences. For example, for the human-orang-utan pair, the divergence is only 2.1% between the two X sequences, but 4.0% between the two Y sequences. In a more detailed analysis, we compute the length of each branch of the four-species tree (Fig. 2). It can be seen that in each branch the Y sequence has evolved faster than the X sequence but the ratio Y/X varies considerably among branches, probably because the number of substitutions in each branch is relatively small. When all branches are considered together, we obtain $Y/X = (14.95 + 2.95 + 1.15 + 1.75 + 2.25)/(5.95 + 1.75 + 0.45 + 1.10 + 1.00) = 2.25$. This is slightly larger than the ratio obtained from synonymous substitutions from the ZFY and ZFX genes^{6–8}: for example, from Fig. 1a of ref. 7, the ratio is $0.17/0.09 = 1.89$ for human ZFY and ZFX.

We now use the Y/X ratio to estimate α_m and to compare the result with that of Miyata *et al.*⁴. Under the assumption that errors in DNA replication during germ-cell division are the primary source of mutation⁴, $\alpha_m \approx \alpha$ (where α is the male-to-female ratio of the number of germ-cell divisions per generation) and the expected mutation ratio of the Y chromosome to autosomes is $Y/A = 2\alpha/(1+\alpha)$ and that of the X chromosome to autosomes is $X/A = (2/3)(2+\alpha)/(1+\alpha)$. As α increases from 1 to infinity, Y/A increases from 1 to 2, while X/A decreases from 1 to $2/3 \approx 0.67$. Miyata *et al.* estimated $Y/A = 2.2$ from the sequences of a Y-linked and an autosomal pseudogene in humans, and $X/A = 0.60$ from the synonymous rates in X-linked and autosomal genes from humans and rodents. As the two estimates exceed the lower and upper limits for the ratios, respectively, they imply that α_m is infinitely large. In contrast, our estimate of $Y/X = 2.3$ leads to $\alpha_m = 6$ from the relation $Y/X = 3\alpha/(\alpha+2)$.

Note that we are actually interested in the ratio $E(Y)/E(X)$, where E means taking expectation, although in practice we use the observed values of X and Y to estimate the ratio, so we are actually estimating Y/X. As $E(Y/X) \approx E(Y)/E(X) + E(X)V(Y)/E(X)^2$, where $V(Y)$ is the variance of Y, $E(Y/X) > E(Y)/E(X)$, and so an observed Y/X tends to overestimate $E(Y)/E(X)$. Despite this fact, all of the Y/X ratios in Fig. 2 are smaller than 3, the value expected for large α . When all branches are pooled together, the overestimation is not serious because the observed Y/X is 2.25 (see above) and $E(Y)/E(X)$ is estimated to be ~2.21. Using the formula $V(Y/X) \approx V(Y)/E(X)^2 + E(Y)^2V(X)/E(X)^4$, we obtain a standard error of ~0.36. Our estimate of $E(Y)/E(X)$ is therefore 2.21 ± 0.36 , from which we find that the 95% confidence interval for α_m is roughly from 2 to 84. As the upper bound is large, we cannot claim that our estimate of α_m is significantly lower than that of Miyata *et al.* We note, however, that those estimates from the ZFY and ZFX coding regions are also smaller than 6.

In this study we have used clearly homologous sequences to avoid possible differences in mutation rates for non-homologous sequences^{5,9}. Of course, as in the case of synonymous changes, introns may be subject to selective constraints. In the present case, however, selection is unlikely to be more effective against the Y intron than against the X intron because ZFY is expressed only in males whereas ZFX is expressed in both males and females.

Another factor that could possibly distort the ratio is sequence exchange such as gene conversion between the X and Y sequences in a species. Although we cannot exclude this possibility, the uniformity in divergence between a Y and an X sequence (about 30% for all pairs; Table 1), and the uniform distribution of nucleotide differences along the intron (data not shown) indicate that there was no exchange between the two sequences

* To whom correspondence should be addressed.

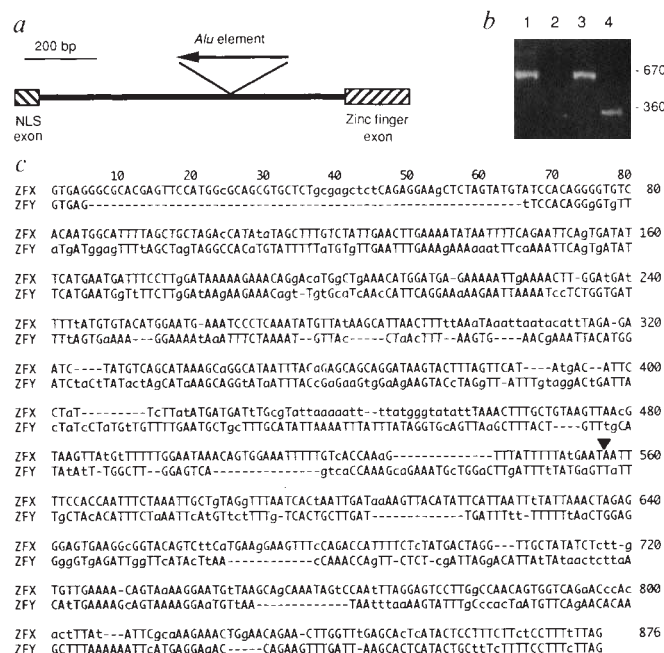
TABLE 1 Number of nucleotide substitutions per 100 sites between intron sequences

Sequences	X intron				Y intron		
	Human	Orang	Baboon	Sq. monkey	Human	Orang	Baboon
Orang X	2.1 ± 0.6						
Baboon X	3.0 ± 0.7	3.5 ± 0.8					
Sq. monkey X	7.6 ± 1.2	7.3 ± 1.1	7.7 ± 1.2				
Human Y	30.6 ± 2.8	31.1 ± 2.8	31.1 ± 2.8	32.2 ± 2.9			
Orang Y	31.6 ± 2.9	32.6 ± 2.9	32.1 ± 2.9	33.3 ± 3.0	4.0 ± 0.8		
Baboon Y	30.2 ± 2.8	31.2 ± 2.9	30.7 ± 2.8	31.6 ± 2.9	6.3 ± 1.1	5.9 ± 1.0	
Sq. monkey Y	32.4 ± 2.9	34.0 ± 3.0	33.2 ± 2.9	32.8 ± 2.9	18.4 ± 2.0	17.8 ± 1.9	17.9 ± 2.0

The mean and standard error are estimated according to ref. 12, a method that takes into account unequal frequencies of the four nucleotides. Gaps are not included in the comparison.

FIG. 1 Structure and sequence alignment of the last intron of primate zinc-finger X (ZFX) and Y (ZFY) genes. *a*, Schematized gene structure of primate genomic DNA fragments containing the last intron of ZFX (including an Alu element inserted in the antisense orientation) and ZFY genes, 65 bp of the upstream putative nuclear localization sequence (NLS) exon and 185 bp of the downstream zinc-finger-domain exon. *b*, Linkage analysis by polymerase chain reaction (PCR) amplification. Squirrel monkey genomic DNA amplified with ZFX- or ZFY-specific primers internal to the intron: lane 1, female, ZFX; lane 2, female, ZFY; lane 3, male, ZFX; lane 4, male, ZFY. Fragment sizes in bp are indicated. *c*, Alignment of the human ZFX and ZFY zinc-finger introns. The alignment excludes the flanking partial exon sequences and the Alu element (the insertion site is indicated by an arrowhead between positions 556 and 557). Capitalized residues indicate nucleotides conserved in ZFX or ZFY in all four species.

METHODS. Primate genomic DNA (*Homo sapiens*, *Pongo pygmaeus*, *Papio cynocephalus* and *Saimiri boliviensis*) was isolated from liver tissue; the X-linked intron was isolated from female DNA, the Y-linked intron from male DNA, except for the squirrel monkey, for which both introns were isolated from male DNA. Available cDNA sequences of mouse (*zfx*, *zfa*, *zfy1*, *zfy2*) and human (ZFX, ZFY) were used to design a pair of degenerate oligonucleotide primers within the NLS domain and the zinc-finger domain (corresponding to nucleotides 1,161–1,186 and 1,441–1,417 in the human ZFX coding sequence¹³). Introns were amplified by PCR reactions containing 10 ng μl^{-1} genomic DNA, 50 mM KCl, 10 mM Tris, pH 8.8, 0.1% Triton X-100, 200 μM dNTP, 2.5 mM MgCl_2 , 1 μM of each primer and 0.04 U μl^{-1} *Taq* DNA polymerase under a thermocycle regime of denaturation, 94 °C for 1 min; annealing, 65 °C for 1 min 20s, and extension, 72 °C for 2 min 10s + 4s autoextension for 30 cycles and a single final extension of 5 min at 72 °C. The intron product DNA (~1,450 bp for ZFX and ~1,030 bp for ZFY) was gel-purified (and if necessary reamplified), cloned into pBluescript SK+ and enzymatically sequenced^{14–16} using vector-specific and 14 internal primers. For each intron, the complete sequence of both strands was determined from two recombinants derived from independent PCR reactions and potential polymorphisms were checked using direct sequencing of gel-purified PCR product¹⁷. Final sequences were aligned using the CLUSTAL¹⁸ program and then optimized by hand. To demonstrate that the ZFY genes are Y-linked,



fragments unique to the ZFX (~670 bp, including Alu) or ZFY (~360 bp) intron homologues were amplified from male and female primate genomic DNA; for each species the ZFY homologue could be amplified from only male DNA, whereas the ZFX homologue could be amplified from both male and female DNA. Genbank accession numbers are ZFX: X58925, X58932, X58930, X58935; ZFY: X58926, X72698, X58931, X58936 for human, orang-utan, baboon and squirrel monkey, respectively.

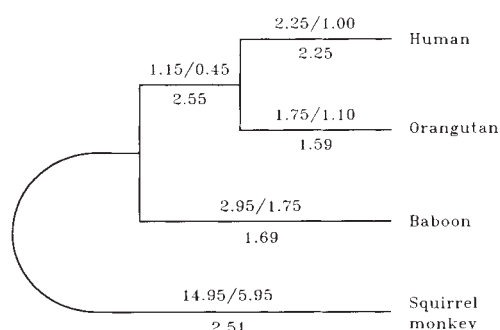


FIG. 2 The numbers above each branch are branch lengths estimated from the Y-linked and the X-linked intron sequences (Table 1), respectively, and the number below each branch is the ratio of the preceding two numbers. The least-squares method was used for the estimation.

in the four lineages. The much closer similarity between human ZFY and ZFX than between human ZFY and mouse *Zfy-1* and *Zfy-2* has been taken as evidence of gene conversion between ZFY and ZFX during primate evolution⁷. However, the hypothesized conversion event would have occurred before the squirrel monkey-human split, for two reasons. First, the divergence (~30%) in intron between every ZFY and ZFX pair is much larger than the 18% divergence between the squirrel monkey and human ZFY (Table 1). Second, the 25% divergence⁷ at synonymous sites between human ZFY and ZFX is also larger than the 18% divergence between the human and squirrel monkey ZFY introns, although smaller than the divergence between human ZFY and ZFX introns (possibly because of stronger selective constraints or stochastic effects).

Actually our estimate of $\alpha_m = 6$ is not far from the estimate of $\alpha_m = 10$ from haemophilia data in refs 1 and 10. But these two haemophilia studies presumably included both indels (inser-

tions and deletions) and point mutations because they did not examine the molecular changes causing the disease in their haemophilia patients. In contrast, our study as well as that of Miyata *et al.*⁴ includes only point mutations. As the mechanism causing point mutation seems to be different from that causing indels, the two types of mutation should be treated separately.

The data obtained in this and previous studies support the view that the mutation rate is much higher in males than in females. However, the male to female mutation rate ratio, α_m , seems to be much lower than Miyata *et al.*'s estimate⁴. This is of interest because whether α_m is very large or relatively small has important implications. A very large α_m implies that errors in DNA replication during germ-cell division is the primary source of mutation and that replication-independent mutagenic factors such as methylation and oxygen radicals are not important. In this case, the effect of generation time on the rate of nucleotide substitution can be strong because the number of germ-cell divisions per unit time is larger in short-living organisms than in long-living ones. On the other hand, if α_m is relatively small, say of the order of only 10, then the contribution of replication-independent mutagenic factors may not be negligible. In this case, the generation-time effect¹¹ may be relatively

weak. The exact magnitude of α_m is therefore a very important issue. □

Received 1 September 1992; accepted 16 February 1993.

- Haldane, J. B. S. *Ann. Eugen.* **13**, 262-271 (1947).
- Penrose, L. S. *Lancet* **ii**, 312 (1955).
- Vogel, F. & Motulsky, A. G. *Human Genetics* (Springer, Berlin and Heidelberg, Germany, 1986).
- Miyata, T., Hayashida, H., Kuma, K., Mitsuyasu, K. & Yasunaga, T. *Cold Spring Harbor Symp. quant. Biol.* **52**, 863-867 (1987).
- Li, W.-H. & Graur, D. *Fundamentals of Molecular Evolution* (Sinauer, Sunderland, MA, 1991).
- Langford, J. & Holland, P. W. H. *J. molec. Evol.* **32**, 310-315 (1991).
- Hayashida, H., Kuma, K. & Miyata, T. *J. molec. Evol.* **35**, 181-183 (1992).
- Pamilo, P. & Bianchi, N. O. *Molec. Biol. Evol.* (in the press).
- Bulmer, M. *Molec. Biol. Evol.* **3**, 322-329 (1986).
- Winter, R. M., Tuddenham, E. G. D., Goldman, E. & Matthews, K. B. *Hum. Genet.* **64**, 156-159 (1983).
- Kohne, D. E. Q. *Rev. Biophys.* **33**, 1-48 (1970).
- Tajima, F. & Nei, M. *Molec. Biol. Evol.* **1**, 269-285 (1984).
- Schneider-Gadick, A. *et al. Nature* **342**, 708-711 (1989).
- Hattori, M. & Sakaki, Y. *Analyt. Biochem.* **152**, 232-238 (1986).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
- Mizusawa, S., Nishimura, S. & Seela, F. *Nucleic Acids Res.* **14**, 1319-1324 (1986).
- Kusukawa, N., Uemori, T., Asada, K. & Kato, I. *BioTechniques* **9**, 66-72 (1990).
- Higgins, D. G. & Sharp, P. M. *Computer Appl. Biosci.* **5**, 151-153 (1989).

ACKNOWLEDGEMENTS. We thank I. Sampaio for squirrel monkey DNA, the Southwest Foundation for Biomedical Research for baboon liver tissues, and M. Shriver for primate DNA, and D. Hewett-Emmett, S.-K. Shyue, A. Zharkikh and Y. Zhong for their help. This study was supported by NIH grants.

Relaxation of imprinted genes in human cancer

Shirley Rainier*, Laura A. Johnson*, Craig J. Dobry*, April J. Ping*, Paul E. Grundy† & Andrew P. Feinberg*‡

* Howard Hughes Medical Institute, and Departments of Internal Medicine and Human Genetics, University of Michigan Medical School, 4520 MSRB I, Ann Arbor, Michigan 48109, USA

† Department of Pediatrics, Cross Cancer Institute and The University of Alberta, 11560 University Avenue, Edmonton, Alberta T6G 1Z2, Canada

GENOMIC imprinting, or parental allele-specific expression of genes, has been demonstrated at the molecular level in insects and mice^{1,2} but not in man. Imprinting as a potential mechanism of human disease is suggested by paternal uniparental disomy of 11p15 in Beckwith-Wiedemann syndrome³ and by maternal uniparental disomy of 15q11-12 in Prader-Willi syndrome⁴. Beckwith-Wiedemann syndrome is characterized by multiorgan overgrowth and predisposition to embryonal tumours such as Wilms' tumour of the kidney⁵. A loss of heterozygosity of 11p15 is also frequently found in a wide variety of tumours, including Wilms' tumour and lung, bladder, ovarian, liver and breast cancers⁶⁻¹¹; 11p15 also directly suppresses tumour growth *in vitro*¹². Two genes in this band, *H19* and insulin-like growth factor-II (*IGF2*) undergo reciprocal imprinting in the mouse, with maternal expression of *H19* (ref. 13) and paternal expression of *IGF2* (ref. 14). Here we find that both of these genes show monoallelic expression in human tissues and, as in mouse, *H19* is expressed from the maternal allele and *IGF2* from the paternal allele. In contrast, 69% of Wilms' tumours not undergoing loss of heterozygosity at 11p showed biallelic expression of one or both genes, suggesting that relaxation or loss of imprinting could represent a new epigenetic mutational mechanism in carcinogenesis.

We have focused on kidney and Wilms' tumours (WT), because both *H19* and *IGF2* are expressed in developing kidney and because WT occurs 700-fold more frequently in Beckwith-Wiedemann syndrome (BWS) than in the general population^{5,15}. DNA from 11 normal kidneys, as well as 17 WTs from these and other patients, were polymorphic for one or more markers. In addition, one rhabdoid tumour and six fetal specimens from various tissues were informative.

TABLE 1 *H19* and *IGF2* alleles expressed in 34 normal and malignant tissues

Tissue	Patient	Stage	<i>H19/RsaI</i>		<i>IGF2/ApaI</i>		<i>IGF2/DR</i>	
			DNA	RNA	DNA	RNA	DNA	RNA
Normal kidney	1		b/b				b/c	b/-
Normal kidney	2		b/b		a/b	a _{pat} /-	a/c	-/c _{pat}
Wilms' tumour	2	I			a/b	a/b	a/c	a/c
Normal kidney	3		a/b	-/b _{mat}	a/b	a/-		
Wilms' tumour	3	II	a/b	-/b _{mat}	a/b	a/b		
Normal kidney	4		a/b	a _{mat} /-				
Wilms' tumour	4	IV	a/b	a _{mat} /-				
Normal kidney	5		a/b	-/b	b/b		b/b	
Wilms' tumour	5	III	a/b	-/b				
Normal kidney	6		a/b	a/-	a/b	-/b _{pat}		
Wilms' tumour	6	III			a/b	a/b		
Normal kidney	7							
Wilms' tumour	7	III			a/b	a/b		
Normal kidney	8		a/b	a/-	a/b			
Wilms' tumour	8	I	a/b	a/-	a/b	a/b		
Normal kidney	9		a/b	a _{mat} /-	a/b	-/b		
Wilms' tumour	9	II			a/b	a/b		
Normal kidney	10		b/b		b/b		b/c	
Wilms' tumour	10	I					b/c	b/c
Normal kidney	11		a/b	-/b				
Wilms' tumour	11	IV	a/b	a/b				
Wilms' tumour	12	II/III	b/b				a/c	a/-
Wilms' tumour	13	II	a/b	a _{mat} /-			b/c	b _{pat} /-
Wilms' tumour	14	I	a/a		a/b	-/b	b/b	
Wilms' tumour	15	I	a/b	a/b	a/b	a/b	a/c	a/c
Wilms' tumour	16	II			b/b	a/b	a/b	a/b
Wilms' tumour	17	I			a/b	a/b		
Rhabdoid tumour	18		b/b		a/b	a/b	b/c	
Fetal liver	19		b/b		a/b	-/b	b/c	-/c
Fetal brain	19		b/b		a/b		b/c	-/c
Fetal kidney	19		b/b		a/b	-/b		
Fetal kidney	20		a/b	a/-				
Fetal heart	20		a/b	a/-				
Fetal lung	20		a/b	a/-				

Seventeen normal tissues and 17 cancers polymorphic for *H19*, *IGF2*, or both, and not showing loss of DNA heterozygosity, were examined for monoallelic or biallelic expression, as described in the legend to Fig. 1. Notation: a/b, both alleles present; a/-, monoallelic expression of a allele; a_{mat}/-, monoallelic expression of maternal a allele; -/b_{pat}, monoallelic expression of paternal b allele.

Monoallelic expression of *H19* in fetal tissues has been described¹⁶, although the parent of origin was unknown. Consistent with that study, *H19* showed monoallelic expression in 15 informative tissues, including seven normal kidneys that had been resected from informative children, five WTs not showing loss of heterozygosity (LOH), and three fetal tissues (Table 1). To determine the parental allele of origin, blood specimens from

‡ To whom correspondence should be addressed.

one or both parents were also examined in 12 of these cases. In all six cases in which parental origin could be determined, the expressed allele was maternal (Fig. 2a; Table 1). *IGF2* also showed monoallelic expression in ten informative tissues (four normal kidneys, three WT, three fetal tissues; Table 1). The allele of origin could be ascertained in three cases and was paternal in all (Fig. 2b; Table 1). Patient 13 directly demonstrated reciprocal imprinting of *H19* and *IGF2* in the same tissues (Fig. 2a, b). Thus, *H19* and *IGF2* displayed imprinting with inactivation of paternal and maternal alleles, respectively.

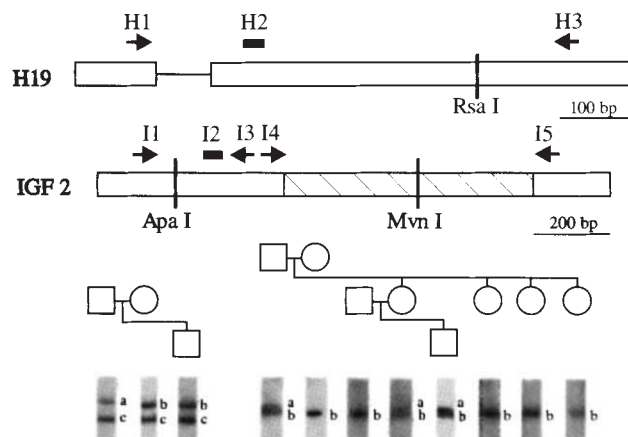


FIG. 1 Analysis of transcribed *H19* and *IGF2* polymorphisms. Maps of *H19* and *IGF2* and three pedigrees exhibiting mendelian segregation of a polymorphic *IGF2* dinucleotide repeat (DR). Exons are denoted by open rectangles, restriction endonuclease cleavage sites by vertical lines, the DR by a hatched rectangle, probes by horizontal bars, and polymerase chain reaction (PCR) primers by arrows.

METHODS. To assess allele-specific expression of *H19*, we took advantage of a transcribed *RsaI* polymorphism by reverse-transcribing RNA and amplifying cDNA by PCR as described¹⁶, but using new primers that permit a single round rather than nested PCR. *IGF2* was analysed using two transcribed polymorphisms, a previously reported *ApaI* variant³⁶, and a newly described dinucleotide repeat in the 3'-untranslated portion of the gene. Poly(A)⁺ RNA was isolated (FastTrack, Invitrogen) and reverse-transcribed (AMV reverse transcriptase; Seikagaku) using primer H3 (5'-TGGATGCTTGAAGGCTGCT-3') for the *H19/RsaI* polymorphism, primer I3 (5'-CCTCTTTGGTCTTACTGGG-3') for the *IGF2/ApaI* polymorphism, or primer I5 (5'-AGAATCTTAGCGGACTTTGGCCT-3') for the *IGF2/DR* polymorphism. *H19* cDNA was amplified by PCR³⁷ using primers H1 (5'-TACAACCACTGCACCTACCTG-3') and H3 in 10 mM Tris, pH 8.3, 25 mM KCl, 1.5 mM MgCl₂, 5 mM NH₄Cl, using a 3-step cycle (94 °C for 30 s, 70 °C for 30 s, and 72 °C for 1 min) for 35 cycles, with a final 10-min incubation at 72 °C. PCR products were ethanol-precipitated, digested with *RsaI*, electrophoresed on a 10% polyacrylamide-15% glycerol gel, and ethidium-bromide stained or electroblotted and hybridized³⁸ with end-labelled probe H2 (5'-CGCTGCTGTTCGATGCTGT-3'). To detect the *IGF2/ApaI* polymorphism, cDNA was amplified with primers I1 (5'-CTTGAGACTTTGAGTCAATTGG-3') and I3 as described³⁶. PCR products were ethanol-precipitated, digested with *ApaI*, electrophoresed on a 10% polyacrylamide-15% glycerol gel, and then stained with ethidium bromide or electroblotted and hybridized with end-labelled probe I2 (5'-CCTCTGACTGCTCTGTGATT-3'). To detect the *IGF2/DR* polymorphism, cDNA was amplified with primers I4 (5'-ACTTTATGCATCCCCGACGTACA-3') and I5. PCR was done in the same buffer as that described for the *H19/RsaI* polymorphism, but using a two-step cycle (94 °C for 30 s, 72 °C for 1 min) for 35 cycles, with a final 10-min incubation at 72 °C. As the DR is relatively large, the polymorphism was detectable by cleaving the PCR product at a central *MvnI* site. Each half of the DR was visualized separately by end-labelling either primer I4 or I5 before PCR, although the left-sided variation occurred more frequently and was used in all the experiments described here. Digests were then electrophoresed on an 8% polyacrylamide-urea gel. *H19* DNA and RNA products were easily distinguished because of the presence of an intervening intron. Duplicate poly(A)⁺ RNA samples were incubated in the presence (+RT) and absence (-RT) of reverse transcriptase, and only those samples showing no amplification in the absence of reverse transcriptase were studied. As an additional control, reverse transcription was done with a tailed primer³⁹ to ensure amplification of RNA only; there was no amplification of DNA in control experiments.

To assess imprinting in cancers, 42 WT were screened for those not exhibiting LOH of 11p15. Surprisingly, of 16 informative tumours retaining heterozygosity, 11 (69%) expressed both maternal and paternal alleles of *H19*, *IGF2*, or both (Table 1; Fig. 2c). The frequency of biallelic expression was greater for *IGF2* (77%) than for *H19* (29%), although this difference was not statistically significant ($P=0.052$). Biallelic expression of *H19* and *IGF2* was concordant in one case (patient 15; Table 1) but was discordant in two (patients 3 and 8; Table 1). Thus, relaxation of an imprint seems to be necessary but not sufficient

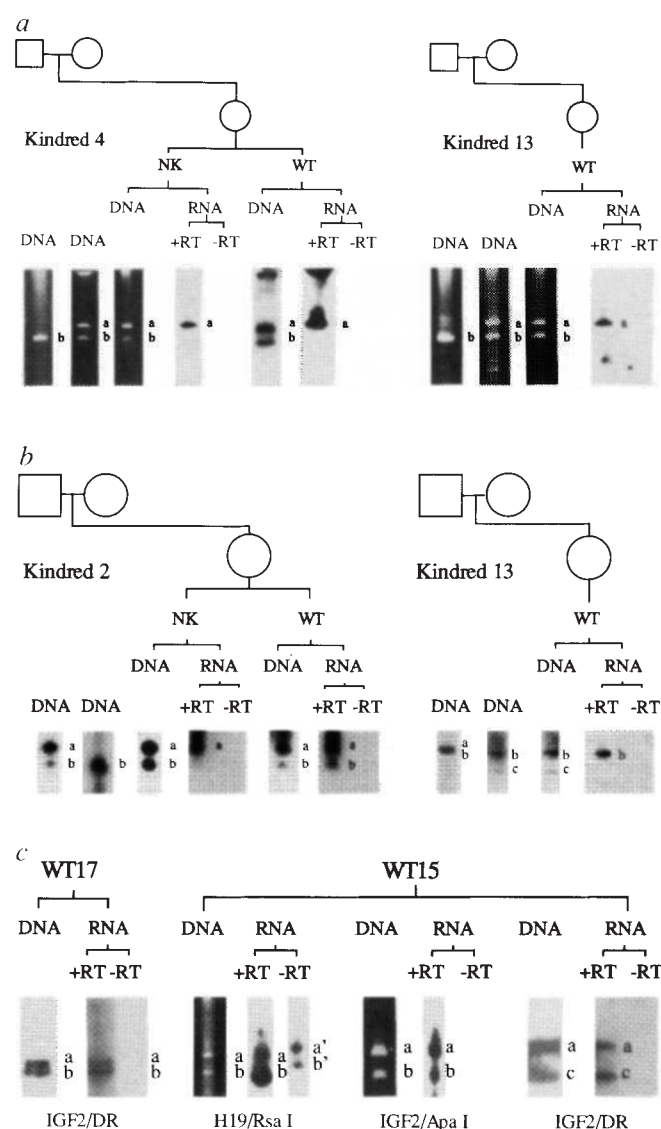


FIG. 2 Imprinting of *H19* and *IGF2* genes. **a**, Maternal monoallelic expression of *H19* in normal kidney (NK) and Wilms' tumour (WT). Kindreds 4 and 13 were analysed using the *RsaI* polymorphism (alleles a and b) as described in the legend to Fig. 1. Both NK and WT of patient 4 and WT of patient 13 show monoallelic expression of the maternal allele. **b**, Paternal monoallelic expression of *IGF2* in normal kidney. Kindred 2 was analysed using the *ApaI* polymorphism, and kindred 13 using the dinucleotide repeat (DR) polymorphism, as described in the legend to Fig. 1. **c**, Biallelic expression of *H19* and *IGF2* in Wilms' tumours. WT17 was analysed using the *IGF2/DR* polymorphism and shows biallelic expression. WT15 was informative for all three polymorphisms. Both *H19* and *IGF2* show biallelic expression, as does the WT from patient 2 (see **b**). This is a qualitative assay for the presence of one or two alleles, and quantitative inferences should not be drawn because of, for example, the presence of heterodimers in PCR products. A single DNA-contaminated RNA sample from patient 15 was deliberately included to illustrate the larger-sized fragments (a', b') resulting from amplification of genomic sequences.

for biallelic expression. These changes were not limited to WT, as a rhabdoid tumour also expressed both *IGF2* alleles.

These experiments lead to two principal conclusions. First, some human genes, like those of other species, undergo allele-specific imprinting. In particular, *IGF2* and *H19* are reciprocally imprinted, with expression of the maternal *H19* and paternal *IGF2* alleles. These data thus lend support to an imprinting model for two other probably distinct but similarly paired genes on 15q11-12 that cause Prader-Willi⁴ and Angelman¹⁷ syndromes, which also involve abnormal (but in those cases decreased) growth. Indeed, genomic imprinting may have evolved because of differing growth demands of maternally and paternally derived gametes¹⁸. Interestingly, bands 11p15 and 15q11-12 map to adjacent regions of chromosome 7 in the mouse¹⁹, which also undergoes imprinting, with overgrowth of paternally disomic chimaeric mice²⁰. Second, the majority of informative non-LOH WTs showed biallelic expression of *IGF2*, *H19*, or both. That these changes play a causal role in malignancy is suggested by (1) their high frequency, which is 2–5-fold greater than that of LOH²¹ and WT1 mutations²² in WT; (2) their early onset, as half were of the earliest stage (Table 1); and (3) both genes are included within the paternally disomic region in BWS³, and uniparental disomy is associated with a >50% incidence of cancer²³.

We therefore propose that loss of imprinting (LOI) in neoplasia represents the functional equivalent of uniparental disomy in BWS. As transgenic mice lacking a functional paternal *IGF2* gene undergo growth retardation¹⁴, *IGF2* has been proposed as a candidate gene for BWS^{24–26}. Overexpression of *IGF2* seen in most WTs²⁷ could be caused by this alteration. It is important to note, however, that *H19* also showed LOI. *H19* is abundantly transcribed in a wide variety of murine tissues but is of cryptic function and may not encode a protein²⁸. *IGF2* and *H19* map to a region of less than 90 kilobases (kb), and thus a regulatory domain lying between them may act as an allele-specific switch controlling the exclusive expression of *IGF2* or *H19* (ref. 29). Thus, biallelic expression of *H19* and *IGF2* in WTs could be caused by a disturbance in this switch, permitting but not requiring expression of both alleles.

As demonstrated in colorectal cancer, multiple genes are involved in multistep neoplastic progression³⁰. We do not believe that the relaxed imprinting seen in these experiments involves the tumour suppressor locus detected by 11p15 LOH^{6–11} because we have mapped this activity to a region distinct from the *H19/IGF2* region¹², and cytogenetic data also indicate that more than one region is involved²⁵. A candidate mechanism for this alteration is DNA hypomethylation, which has been reported in a wide variety of premalignant and malignant tumours^{31–33}, but for which no gene has yet been shown to be specifically activated in cancer. Furthermore, 1% of normal children under one year old harbour WT precursor lesions that spontaneously regress^{34,35}. It is difficult to imagine how a process of such high frequency could occur by conventional mutational mechanisms. Regardless of the precise role that this process plays in neoplasia, these data provide evidence for a novel epigenetic mutational mechanism in cancer, namely relaxation of genomic imprinting. □

14. DeChiara, T. M., Robertson, E. J. & Efstratiadis, A. *Cell* **64**, 849–859 (1991).
15. Wiedemann, H. R. *Eur. J. Pediatr.* **141**, 129 (1983).
16. Zhang, Y. & Tycko, B. *Nature Genet.* **1**, 40–44 (1992).
17. Wagstaff, J. *et al. Nature Genet.* **1**, 291–294 (1992).
18. Haig, D. & Graham, C. *Cell* **64**, 1045–1046 (1991).
19. Davison, M. T. *et al. Cytogenet. Cell Genet.* **58**, 1152–1189 (1991).
20. Ferguson-Smith, A. C., Cattane, B. M., Barton, S. C., Beechey, C. V. & Surani, M. A. *Nature* **351**, 667–670 (1991).
21. Wadey, R. B. *et al. Oncogene* **5**, 901–907 (1990).
22. Little, M. H. *et al. Proc. natn. Acad. Sci. U.S.A.* **89**, 4791–4795 (1992).
23. Junien, C. *Curr. Opin. Genet. Dev.* **2**, 431–438 (1992).
24. Little, M., van Heyningen, V. & Hastie, N. *Nature* **351**, 609–610 (1991).
25. Hoovers, J. M. N., Dietrich, A. J. J. & Mannens, M. M. A. *Lancet* **339**, 1228 (1992).
26. Mannens, M. *The Molecular Genetics of Wilms' Tumour and Associated Congenital Diseases* (Univ. Amsterdam Press, Amsterdam, 1991).
27. Reeve, A. E., Eccles, M. R., Wilkins, R. J., Bell, G. I. & Millow, L. J. *Nature* **317**, 258–260 (1985).
28. Brannan, C. I., Dees, E. C., Ingram, R. S. & Tilghman, S. M. *Molec. cell. Biol.* **10**, 28–36 (1990).
29. Zemel, S., Bartolomei, M. S. & Tilghman, S. M. *Nature Genet.* **2**, 61–65 (1992).
30. Fearon, E. P. & Vogelstein, B. *Cell* **61**, 759–767 (1990).
31. Feinberg, A. P. & Vogelstein, B. *Nature* **301**, 89–92 (1983).
32. Jones, P. A. *Cancer Res.* **46**, 461–466 (1986).
33. Goetz, S. E., Vogelstein, B., Hamilton, S. R. & Feinberg, A. P. *Science* **228**, 187–190 (1985).
34. Beckwith, J. B., Kiviat, N. B. & Bonadio, J. F. *Embryol. Dev.* 1–36 (1990).
35. Bennington, J. L. & Beckwith, J. B. *Atlas of Tumor Pathology* (Armed Forces Inst. Pathology, Washington, DC, 1975).
36. Tadokoro, K., Fujii, H., Inoue, T. & Yamada, M. *Nucleic Acids Res.* **19**, 6967 (1991).
37. Saiki, R. K. *et al. Science* **230**, 1350–1354 (1985).
38. Church, G. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1991–1995 (1984).
39. Shuldiner, A. R., Tanner, K., Moore, C. A. & Roth, J. *BioTechniques* **11**, 760–763 (1991).

ACKNOWLEDGEMENTS. We thank The Childrens Cancer Group and Pediatric Oncology Group for specimens, J. Bonadio for pathological review, J. Lowe, D. Law, G. Nabel and T. Glover for discussion, and A. Huntzicker for preparing the manuscript. This work was supported by the National Institutes of Health.

Relaxation of insulin-like growth factor II gene imprinting implicated in Wilms' tumour

Osamu Ogawa, Michael R. Eccles, Jenny Szeto, Leslie A. McNoe, Kankatsu Yun*, Marion A. Maw, Peter J. Smith† & Anthony E. Reeve‡

Cancer Genetics Laboratory, Department of Biochemistry and

*Department of Pathology, University of Otago, Dunedin, New Zealand

†Department of Pathology, University of Queensland, Herston, Queensland 4006, Australia

GENOMIC imprinting has been implicated in the onset of several embryonal tumours but the mechanism is not well understood^{1–3}. Maternal chromosome 11p15 loss of heterozygosity⁴ and paternal chromosome 11 isodisomy^{5,6} suggest that imprinted genes are involved in the onset of Wilms' tumour and the Beckwith-Wiedemann syndrome. The insulin-like growth factor II (*IGF2*) gene located at 11p15.5 has been put forward as a candidate gene as it is maternally imprinted (paternally expressed) in the mouse⁷, and is expressed at high levels in Wilms' tumours^{8,9}. We report here that the *IGF2* gene is expressed from the paternal allele in human fetal tissue, but that in Wilms' tumour expression can occur biallelically. These results provide, to our knowledge, the first evidence that relaxation of imprinting may play a role in the onset of disease and suggest a new genetic mechanism involved in the development of cancer.

The allele specificity of *IGF2* expression in human fetal kidneys was determined using an *ApaI*/restriction enzyme fragment length polymorphism (RFLP)¹⁰ in the 3' untranslated portion of the *IGF2* gene (Fig. 1). DNA and RNA from six fetal kidneys (12-week gestation) were amplified using polymerase chain reaction (PCR) and the products digested with *HinfI*/*ApaI*. In all fetal kidneys *IGF2* RNA was monoallelically expressed; two representative cases are shown in Fig. 1. In three additional cases, fetal tissue and maternal blood were used to determine the parental origin of *IGF2* expression, and Fig. 2

Received 23 November 1992; accepted 22 February 1993.

1. Willison, K. *Trends Genet.* **7**, 107–109 (1991).
2. Holliday, R. *Development* suppl. 125–129 (1990).
3. Henry, I. *et al. Nature* **351**, 665–667 (1991).
4. Nicholls, R. D., Knoll, J. H. M., Butler, M. G., Karam, S. & Lalonde, N. *Nature* **342**, 281–285 (1989).
5. Pettenati, M. J. *et al. Hum. Genet.* **74**, 143–154 (1986).
6. Mannens, M. *et al. Hum. Genet.* **81**, 41–48 (1988).
7. Reeve, A. E., Sih, S. A., Raizis, A. M. & Feinberg, A. P. *Molec. cell. Biol.* **9**, 1799–1803 (1989).
8. Skinner, M. A., Vollmer, R., Huper, G., Abbott, P. & Ingelhart, J. D. *Cancer Res.* **50**, 2303–2306 (1990).
9. Ali, I. U., Lidereau, R., Theillet, C. & Callahan, R. *Science* **238**, 185–188 (1987).
10. Fujimori, M. *et al. Cancer Res.* **51**, 89–93 (1991).
11. Ehlen, T. & Dubeau, L. *Oncogene* **5**, 219–223 (1990).
12. Koi, M. *et al. Science* (in the press).
13. Bartolomei, M., Zemel, S. & Tilghman, S. M. *Nature* **351**, 153–155 (1991).

‡ To whom correspondence should be addressed.

TABLE 1 Comparison of heterozygosity status at *IGF2* locus and allelic transcription in Wilms' tumours

Tumour	<i>IGF2</i> LOH	<i>IGF2</i> RNA
1	LOH	M (a)
2	LOH	M (b)
3	N	M (a)
4	N	M (b)
5	N	B
6	N	B
7	N	B
8	N	B

Eight of 30 patients used in this study were heterozygous for the *IGF2* *Apa*I RFLP. Tumour loss of heterozygosity (LOH), no loss of heterozygosity (N), and whether *IGF2* transcription is monoallelic (M) or biallelic (B), are indicated. Expressed *IGF2* alleles are designated (a) and (b), as defined in Fig. 1 legend.

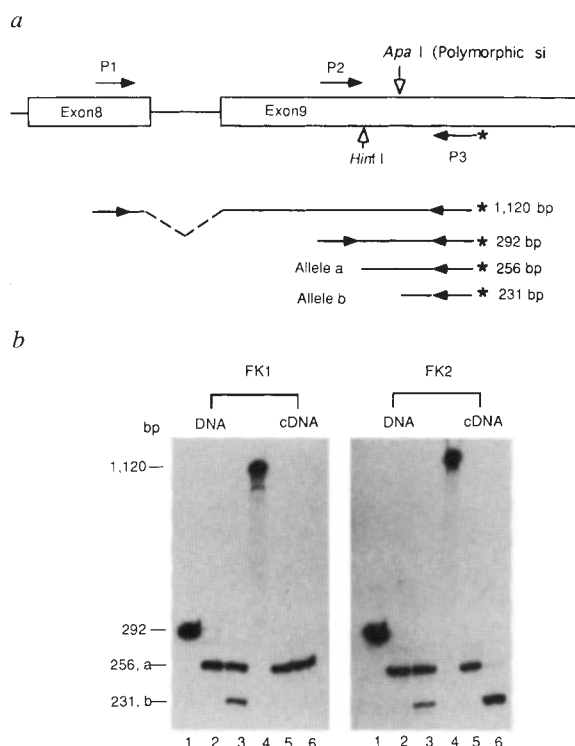


FIG. 1 *a*, Map of the 3' end of the *IGF2* gene and size of PCR products. cDNA was produced from total RNA by reverse transcription using the P3 primer²¹. To eliminate DNA contamination during RNA-PCR, the cDNA was amplified using primer P1 and primer P3 (end-labelled with ³²P), which encompassed the last intron of the *IGF2* gene. Following cDNA-PCR, the RNA-specific product (1,120 bp) was purified from the DNA-specific product (1,400 bp) by agarose gel electrophoresis, digested with *Hinf*I and *Apa*I, and electrophoresed on 6% denaturing polyacrylamide gels which were then autoradiographed. For DNA-PCR, P2 and P3 (³²P-end-labelled) primers were used. PCR amplifications were done using an initial denaturation for 4 min at 94°C, then 35 cycles at 94°C for 1 min, 55°C for 2 min, 72°C for 3 min, followed by a final extension at 72°C for 3 min. The *Apa*I-undigested and digested alleles were designated allele a and allele b, respectively. Three primers used were: P1 (5'-TCCTGGAGACGTACTGTGCTA-3'), P2 (5'-CTTGGACCTTGAGTCAAATTGG-3') and P3 (5'-GGTCGTGCCAATTACATTCA-3'). *b*, Monoallelic expression observed in two fetal kidneys (FK1 and FK2). Lanes 1–3 and 4–6 show DNA-PCR and cDNA-PCR products, respectively. Lanes 1 and 4 contain undigested PCR products, lanes 2 and 5 show PCR products digested with *Hinf*I, and lanes 3 and 6 show PCR products digested with *Hinf*I and *Apa*I. Under these PCR conditions, the ratio of the two intensities of the polymorphic alleles in heterozygous normal DNA was constant at ~3:1 (see lane 3 and Figs 2 and 3), which was considered to represent equal amounts of the two polymorphic alleles, as a result of heterodimer formation which was refractory to *Apa*I digestion.

shows that in fetal muscle and kidney (17-week) *IGF2* is selectively expressed from the paternally inherited allele. Maternal imprinting of human *IGF2* is in agreement with results from mice demonstrating paternal transmission of dwarfism in animals carrying an *IGF2* gene disrupted by targeting¹¹.

We also studied the allele specificity of *IGF2* expression in Wilms' tumours. RFLP experiments have demonstrated that chromosome 11 allele loss in Wilms' tumour is specific for the maternally inherited chromosome 11 (ref. 4). Consequently, in tumours that have undergone 11p loss of heterozygosity (LOH) *IGF2* transcription would be expected to originate from the remaining paternal allele. This is demonstrated in Fig. 3a, where the RNA of two tumours with chromosome 11p LOH was amplified by RNA-PCR and analysed using the *IGF2* *Apa*I RFLP. In both cases paternal monoallelic expression was observed. The specificity of *IGF2* allele expression was then determined in tumours that had retained heterozygosity. Of thirty Wilms' tumours analysed, six were heterozygous at the *IGF2* RFLP site. Two of these heterozygous tumours expressed *IGF2* RNA monoallelically (Fig. 3b), whereas biallelic *IGF2* RNA synthesis was found in the four remaining Wilms' tumours, two examples of which are shown in Fig. 3c. These findings are summarized in Table 1 and indicate that imprinting of the *IGF2* gene in some Wilms' tumours differs from that of fetal kidney from which the tumours originate.

Biallelic *IGF2* expression could indicate that relaxation of *IGF2* imprinting occurs as a somatic event that is involved in tumorigenesis, or alternatively, that Wilms' tumours arise in the fetal kidney from a minor population of embryonal cells that normally express *IGF2* biallelically. From our data it is not possible to resolve whether the relaxation of *IGF2* imprinting is directly related to the onset of Wilms' tumour. Although it could be that the relaxation of *IGF2* imprinting is unrelated to tumorigenesis, the *IGF2* gene has been considered as a candidate gene for tumour onset^{8,9,12}. An increase in *IGF2* expression, resulting from the reduplication of the paternally inherited 11p allele, has been proposed to be important in tumorigenesis^{5,12}. Consequently, whether the expression of *IGF2* messenger RNA is increased in Wilms' tumours as a result of an increase in gene dosage or biallelic expression, the outcome should be identical. Also, biallelic expression could result in an altered timing of transcription from the paternal or maternal alleles and consequently in an altered pattern of *IGF2* synthesis during the cell cycle. Not all heterozygous Wilms' tumours (2/8) expressed *IGF2* biallelically, which was not surprising as several genes are known to be involved in Wilms' tumorigenesis.

IGF2 has also been considered as a candidate gene for the Beckwith-Wiedemann syndrome (BWS), which predisposes to embryonal malignancies and is associated with overgrowth^{9,13–15}. It has been proposed that constitutional paternal isodisomy of chromosome 11 found in some BWS patients^{5,6} may lead to a twofold increase in gene dosage of the active *IGF2* allele¹². Furthermore, in sporadic BWS patients the frequency of constitutional homozygosity of chromosome 11p RFLP markers is highest at the insulin-*IGF2* locus⁵. The potential role of *IGF2* gene dosage in BWS is supported by the observation that chimaeric mouse embryos isodisomic for the distal region of chromosome 7 (syntenic with human chromosome 11p15) show a striking growth enhancement¹⁶.

The observation that *IGF2* imprinting is altered in some Wilms' tumours indicates that this event might also occur constitutionally in some BWS patients¹⁵. Defective germ line imprinting of the *IGF2* gene provides an explanation for the preferential maternal transmission of BWS^{17–20}. In this case, defective imprinting of the inherited maternal *IGF2* allele could lead to constitutional biallelic *IGF2* expression, which would be functionally similar to the paternal chromosome 11p duplications and isodisomies seen in some patients with this syndrome^{5,6}. This mechanism does not necessarily require, however, that maternal imprinting be completely defective in BWS.

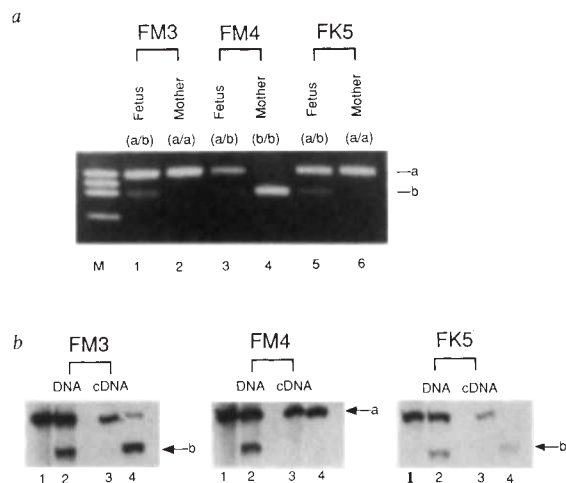


FIG. 2 Parental origin of *IGF2* allele expression in fetal tissues. *a*, Agarose gel electrophoresis of *Apal*-digested DNA-PCR products from three different fetuses and corresponding maternal DNA are shown. All fetuses were heterozygous (lanes 1, 3 and 5), and all mothers were homozygous (lanes 2, 4 and 6) at the *Apal* site. Lane M shows pUC 18/*HaeIII*-digested size markers. *b*, *IGF2* paternal allele expression in fetal tissues. cDNA was analysed in two cases of fetal muscle (FM3 and FM4) and one case of fetal kidney. Lanes 1 and 3 show *HinfI* digest, and lanes 2 and 4 show *HinfI*/*Apal* digest of PCR products. PCR conditions are described in Fig. 1 legend.

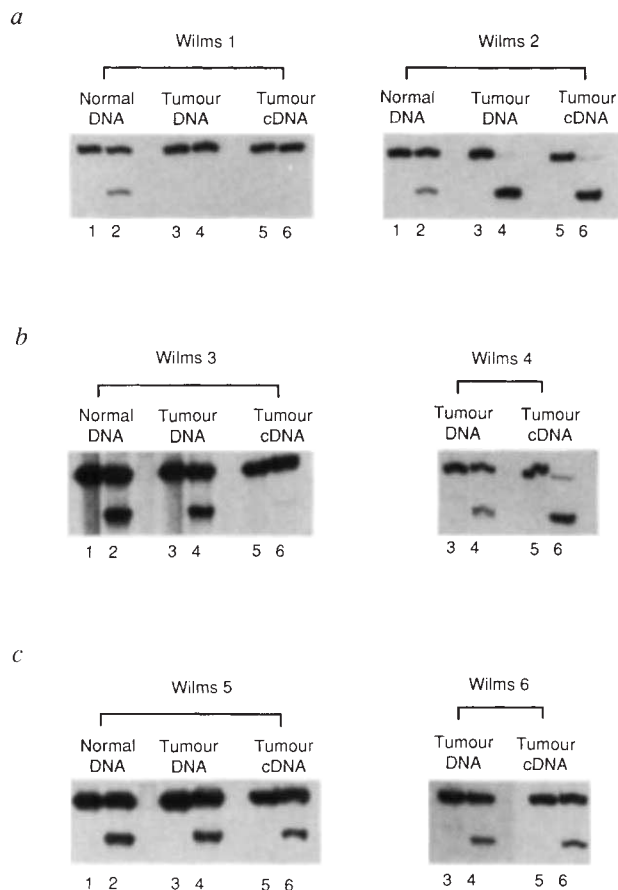


FIG. 3 Analysis of *IGF2* allelic expression in Wilms' tumours. *a*, Tumours with loss of heterozygosity at *IGF2* showing monoallelic *IGF2* RNA synthesis. *b*, Heterozygous tumours with monoallelic *IGF2* transcription. *c*, Heterozygous tumours with biallelic *IGF2* transcription. PCR products were digested with *HinfI* (lanes 1, 3 and 5) and *HinfI*/*Apal* (lanes 2, 4 and 6). PCR conditions are described in Fig. 1 legend.

Accordingly, the magnitude and pattern of growth abnormalities and the risk of tumour development could be determined in some cases by the extent to which imprinting of the *IGF2* maternal allele is defective.

The finding that the normal pattern of *IGF2* imprinting is altered in Wilms' tumours could have wider implications for tumorigenesis. Only a few genes are at present known to be imprinted, and these play a critical role in growth and development. Other growth regulatory genes may also be imprinted and the relaxation of imprinting could be involved in growth disorders and the development of cancer. □

Received 20 January; accepted 10 March 1993.

- Wilkins, R. J. *The Lancet* **1**, 329-331 (1988).
- Reik, W. & Surani, M. A. *Nature* **338**, 112-113 (1989).
- Ponder, B. *Nature* **340**, 264 (1989).
- Junien, C., Van Heyningen, V., Evans, G., Little, P. & Mannens, M. *Genomics* **12**, 620-625 (1992).
- Henry, I. *et al.* *Nature* **351**, 665-667 (1991).
- Grundy, P. *et al.* *The Lancet* **338**, 1079-1080 (1991).
- DeChiara, M., Robertson, E. J. & Efstratiadis, A. *Cell* **64**, 849-859 (1991).
- Reeve, A. E., Eccles, M. R., Wilkins, R. J., Bell, G. I. & Millow, L. J. *Nature* **317**, 258-260 (1985).
- Scott, J. *et al.* *Nature* **317**, 261-262 (1985).
- Tadokoro, K., Fujii, H., Inoue, T. & Yamada, M. *Nucleic Acids Res.* **19**, 6967 (1991).
- DeChiara, T. M., Efstratiadis, A. & Robertson, E. J. *Nature* **345**, 78-80 (1990).
- Little, M., van Heyningen, V. & Hastie, N. *Nature* **351**, 609-610 (1991).
- Olshan, A. F. *Cancer Genet. Cytogenet.* **21**, 303-307 (1986).
- Reik, W. *Trends Genet.* **5**, 331-336 (1989).
- Moore, T. & Haig, D. *Trends Genet.* **7**, 45-49 (1991).
- Ferguson-Smith, A. C., Cattanaach, B. M., Barton, S. C., Beechey, C. V. & Surani, M. A. *Nature* **351**, 667-670 (1991).
- Koufos, A. *et al.* *Am. J. Hum. Genet.* **44**, 711-719 (1989).
- Aleck, K. A. & Hadro, T. A. *Am. J. med. Genet.* **33**, 155-160 (1989).
- Moutou, C., Junien, C., Henry, I. & Bonaiti-Pellie, C. *J. Med. Genet.* **29**, 217-220 (1992).
- Vijoen, D. & Ramesar, R. *J. med. Genet.* **29**, 221-225 (1992).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* 2nd edn (Cold Spring Harbor Laboratory Press, New York, 1989).

ACKNOWLEDGMENTS. We thank R. J. M. Gardner, I. M. Morison and R. J. Wilkins for comments on the manuscript; and G. Abbott, D. Becroft, E. Chan, P. Crossen, B. Delahunt, T. Fiddes, P. Fitzgerald, W. Gillett, N. Gopinathan, G. Grubb, R. Hill, D. Holdaway, M. Kennedy, M. Lewis, R. Massey, D. Mauger, C. McRae, R. Parfitt, R. Shaw and D. Wilson for material used in this study. This work was supported by the Cancer Society of New Zealand, New Zealand Lottery Grants Board, and the Queensland Cancer Trust.

Parental-origin-specific epigenetic modification of the mouse *H19* gene

Anne C. Ferguson-Smith, Hiroyuki Sasaki, Bruce M. Cattanaach* & M. Azim Surani

Wellcome/CRC Institute of Cancer and Developmental Biology, Tennis Court Road, Cambridge CB2 1QR, UK, and Physiological Laboratory, University of Cambridge

*MRC Radiobiology Unit, Chilton, Didcot, Oxford OX1 0RD, UK

THE *H19* gene produces an abundant developmentally regulated transcript of unknown function in normal embryos¹. In the mouse it lies on chromosome 7 and is subject to transcriptional regulation by parental imprinting, which results in the maternally inherited gene being expressed and the paternally inherited gene being repressed². Embryos carrying maternal duplication/paternal deficiency for distal chromosome 7 (MatDi7)³ therefore express a double dose of *H19*. Here we examine the parental-origin-specific epigenetic modifications that may be involved in this regulation by comparing CpG methylation and nuclease sensitivity of chromatin in MatDi7 embryos with normal littermates. We show that specific sites in the CpG island promoter and 5' portion of the gene are methylated only on the paternal allele. Furthermore, active maternal alleles in chromatin of MatDi7 embryos are more sensitive and accessible to nucleases. Therefore hypermethylation and chromatin compaction in the region of the *H19* promoter is associated with repression of the paternally inherited copy of the gene. Most, but not all, of these sites are unmethylated in sperm,

with methylation of the paternal promoter occurring after fertilization. These results contrast with our findings for the closely linked and reciprocally imprinted gene encoding insulin-like growth factor II (ref. 4).

Although the sites of *H19* expression are the same in normal and MatDi7 embryos (*in situ* data; results not shown), the level of transcription in MatDi7 embryos is roughly twice that in normal embryos (Fig. 1a and b). Nuclear run-on analysis (Fig. 1b) shows that the gene is controlled by imprinting mainly at the level of transcription.

MatDi7 embryos die at around day 16 (d16) of gestation and are growth-retarded⁵. The reciprocal PatDi7 (paternal duplication/maternal deficiency) embryos have yet to be identified and die at ~d6–d10 of gestation^{3,5}. It is likely that repression of both alleles of the insulin-like growth factor II gene (*Igf2*) is responsible for the growth retardation but not the inviability of MatDi7 embryos^{3,6}. The increased dosage of *H19* in these embryos could be involved in the lethality because transgenic embryos expressing exogenous *H19* die earlier than d14 (ref. 7).

The control of genetic imprinting is likely to involve heritable modification to DNA and/or chromatin. In this analysis we have looked for modifications associated with the imprinting of *H19* by comparing normal material with that from MatDi7 embryos isolated before d16. Any differences between the embryos could be attributed to the repressed paternal allele absent in MatDi7 embryos in which both alleles are active.

One form of heritable epigenetic modification associated with gene inactivity is CpG methylation, particularly of CpG islands in gene promoters (see later)^{8,9}. The genome undergoes a process of stage and sequence-specific CpG methylation changes in germ cells and in the developing embryo^{10,11}. Moreover, homozygous mutant mice for the enzyme DNA methyltransferase die around mid-gestation, suggesting a vital role for DNA methylation by this enzyme during embryogenesis¹². A role for methylation in imprinting is suggested by analysis of the monoallelic activity of X-linked genes in female mammals^{13,14} and the parental origin-specific DNA methylation of transgenes¹⁵. Gene inactivity is also often accompanied by reduction in the accessibility of chromatin to nucleases such as *MspI* and DNase I (refs 16–18).

Methylation-sensitive restriction enzymes were used to compare the extent of CpG methylation at sites that we mapped in the *H19* locus. Figure 2a shows a restriction map of the gene,

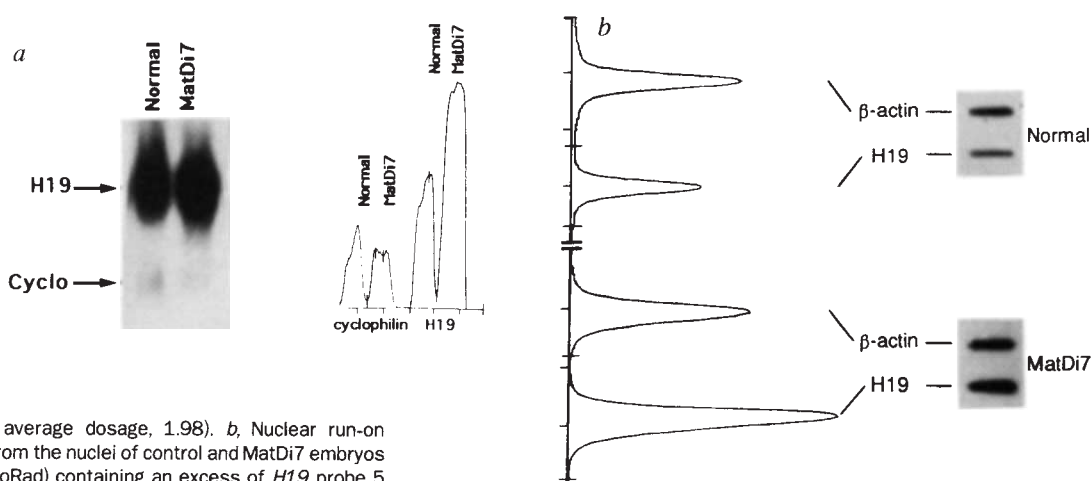
and positions of probes used. The promoter region¹⁹ is CpG-rich and contains a cluster of methylation-sensitive restriction sites. The distribution of sites investigated is shown in Fig. 2b. Eight of the 20 sites analysed reveal methylation of the paternal allele but no methylation of the maternal allele (Fig. 3). Of the four sites available for analysis in the promoter itself, all are differentially methylated, as indicated by the complete digestion of informative high-molecular-weight bands in MatDi7 embryos (Fig. 3a, b and c) and about 50 per cent digestion in normal embryos. This differential methylation in the promoter extends into two *HpaII* sites in the first exon of the gene (Fig. 3a). Thereafter, normal and MatDi7 embryos exhibit the same pattern of DNA methylation, notably at the 3' end around the enhancers (Fig. 3c). Indeed, several sites in the body of the gene show complete methylation of both alleles (data not shown). About 1 kilobase (kb) upstream of the promoter, two sites (*HhaI* and *HaeII*) show differential methylation (Fig. 3b). A summary of the methylation pattern at all sites analysed is shown in Fig. 2c. It illustrates the differential methylation at the maternal and paternal alleles. This study is limited to recognition sites for methylation-sensitive restriction enzymes and we cannot rule out the presence of other differentially methylated CpG dinucleotides.

It is not known when the initial parental-origin-specific signals are established. It is assumed that they occur in the germ line sometime during gametogenesis, and subsequent modifications may then occur post-fertilization in response to these signals²⁰. We have analysed DNA from mature sperm to assess whether methylation of the paternal allele is established before fertilization. Sperm DNA is unmethylated at all sites in the promoter (Fig. 3a and b), in contrast to the paternal allele in normal embryos. As DNA from embryonic and extraembryonic tissues is methylated at d9 of gestation (Fig. 3a), this methylation must occur beforehand. Thus, the paternal promoter becomes methylated after fertilization. Intriguingly, however, the two sites upstream of the promoter are methylated in sperm, in contrast to all other differentially methylated sites (Fig. 3b, asterisk). The single *HaeII* site is uncleaved in sperm; the cleavage with *HhaI* is due to the absence of methylation in the promoter (1.8-kb product), with the upstream site being only very slightly digested (1 kb). The extent and significance of this modification in sperm still needs to be determined.

The distribution of CpG dinucleotides in *H19* was analysed

FIG. 1 MatDi7 embryos express a double dose of *H19*. a, Northern blot analysis of total RNA from MatDi7 embryos and control littermates hybridized to a *H19* cDNA probe (probe 5). The ubiquitously expressed housekeeping gene, cyclophilin D³⁰ (cyclo), was used as a qualitative and quantitative control. Densitometric analysis (right) normalized to control probes showed a ~2-fold increase in *H19* transcript compared with normal littermates ($n = 5$; gestational age, d13–d16; average dosage, 1.98). b, Nuclear run-on quantitation. ³²P-labelled RNA from the nuclei of control and MatDi7 embryos was hybridized to slot blots (BioRad) containing an excess of *H19* probe 5 or β -actin (purified insert cDNA) as a control showing again, that MatDi7 embryos express twice the amount of RNA compared to controls. Thus, imprinting of *H19* is at the level of transcription. These data were reproducible for three embryos; reactions in the presence of α -amanitin were negative (results not shown).

METHODS. Northern analysis was carried out as before³. Densitometry of band intensities was done on a Chromoscan 3 (Joyce-Loebl). Isolation of nuclei, *in vitro* transcription, RNA isolation and quantitation, were all as



described^{4,31}. MatDi7 embryos are duplicated for the region of chromosome 7 distal to the T(7;15)9H translocation breakpoint⁵. They are generated as progeny of the translocation heterozygote intercross T9Hc/+c × T9H+/+, with the female being homozygous for the recessive mutation albino (denoted as c), also distal to T9H. MatDi7 embryos are recognized by the absence of eye pigment. The translocation is on a BALB/c background. Day 1 of gestation (d1) is considered as the morning of vaginal plug identification.

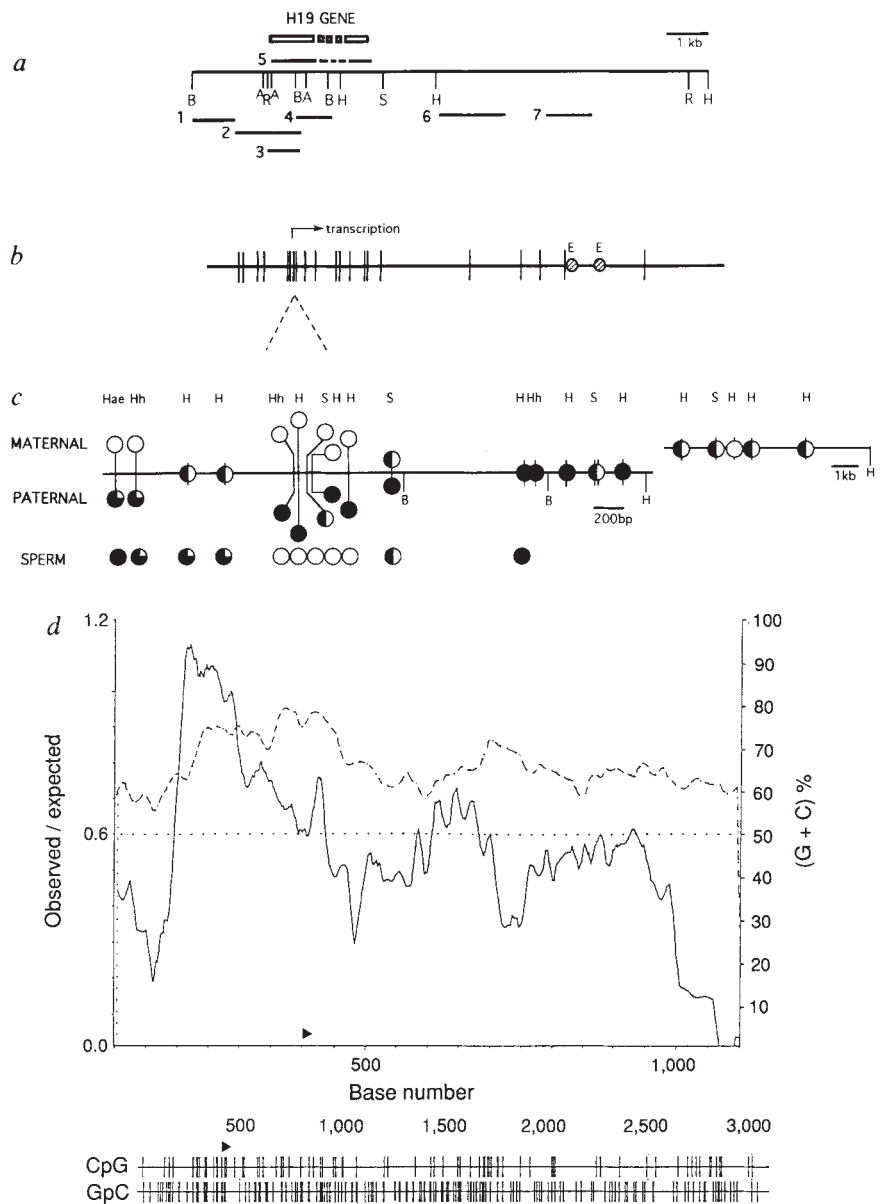
and the 5' region encompassing the promoter was found to be a CpG island (Fig. 2d). CpG islands are found in housekeeping and some tissue-specific genes, and are generally coincident with promoters; their methylation is associated with gene inactivity²¹. Furthermore, CpG islands are never methylated in sperm²¹, just as we have found with *H19*. On the inactive X chromosome, CpG islands become methylated in early embryogenesis, with those on the active X remaining unmethylated²². Thus the differential methylation of *H19* shares similarities with X-linked genes. Methylation in promoter regions has been proposed to affect transcriptional inactivity in a variety of ways. Such modification might silence the paternal allele by indirectly preventing active transcription complex formation²³⁻²⁶, or by directly inhibiting binding of particular transcription factors^{27,28}.

To examine the epigenetic state of the 5' region of the *H19* gene further, we compared the sensitivity of chromatin to nucleases in MatDi7 and control embryos. Accessibility of the promoter to *MspI* is greater in MatDi7 nuclei (Fig. 4), as indicated by the more rapid disappearance of the 2.7-kb *BamHI* fragment and the more rapid appearance of the 0.7-kb digestion product. The 1-kb internal fragment did not show increased sensitivity. The same result was obtained using *EcoRI*

(not shown). With DNase I, we identified a single hypersensitive site 700 base pairs (bp) upstream of the first internal *BamHI* site, in the promoter itself, in addition to the five hypersensitive sites previously identified in the downstream enhancer region⁷. Although there was no difference in the sensitivity of this 3' portion, the 5' site showed greater sensitivity on the maternally derived chromosome. This allele-specific difference was confirmed in nuclei from *Mus musculus* × *Mus spretus* intercrosses, in which a DNA polymorphism indicated the maternal origin of the more sensitive maternal allele (H.S. and W. Reik, unpublished results). Hence the presence of inactive paternally inherited *H19* chromatin is associated with a decrease in nuclease accessibility, and this resistance to nucleases can be correlated with the increase in DNA methylation in the 5' region of the gene.

In conclusion, increased methylation and chromatin compaction is consistent with repression of the paternally inherited *H19* gene. As methylation of the *H19* promoter occurs after fertilization and is not inherited from sperm, the molecular characterization of the initial imprint remains a key objective. The imprinting of *H19* can also be considered in the context of a larger domain containing the reciprocally imprinted *Igf2* gene located

FIG. 2 Schematic representation of the mouse *H19* locus. *a*, Restriction map illustrating the intron-exon structure of *H19*. Probes 1–7 are indicated. Restriction enzyme digestion sites are: H, *HindIII*; B, *BamHI*; S, *Sall*; R, *EcoRI*; A, *Apal* (only 5' *Apal* sites are shown). *b*, Distribution of the methylation-sensitive restriction enzyme sites in the *H19* locus. The start of transcription and the position of the two downstream enhancers (E) are indicated¹⁹. The methylation status of all sites were analysed by Southern hybridization using the probes shown in *a*. *c*, Summary of methylation of the maternally (upper circles) and paternally (lower circles) inherited alleles of *H19*. Open circles indicate no methylation at the site, and filled circles that the site is methylated. Half-filled circles represent partial methylation, and three-quarter-filled circles indicate sites that are only slightly digested with methylation-sensitive enzymes. Sites not showing differential methylation are represented as positioned on the line. The status of the differentially methylated sites was analysed in sperm DNA and is shown as circles beneath the paternal allele. These data are consistent for several embryos, tissues and gels analysed. Restriction sites: H, *HpaII*; S, *SmaI*; Hh, *HhaI*; Hae, *HaeIII*. *d*, Analysis of the distribution of CpG dinucleotides in *H19*. The (G+C) percentage (dotted line) and the observed/expected CpG ratio (smooth curve) for 1 kb at the 5' end of the gene is plotted as described in ref. 32. The locations of individual CpG and GpC dinucleotides for the whole gene are shown beneath the graph; the region analysed in the graph is underlined. Arrowheads indicate the start site of transcription. Analysis of nucleotides –200 to +300 (relative to the arrowhead) indicates (G+C)=61.8% with observed/expected CpG=0.63. We can thus classify the region here as a CpG island according to the criteria defined in ref. 32. The CpG-rich region is concentrated at –170 to +90 (obs/exp=0.81). Sequences were analysed using the GCG package and programs CpGPLOT and MAPLOT and sequences derived from our laboratory and the EMBL database.



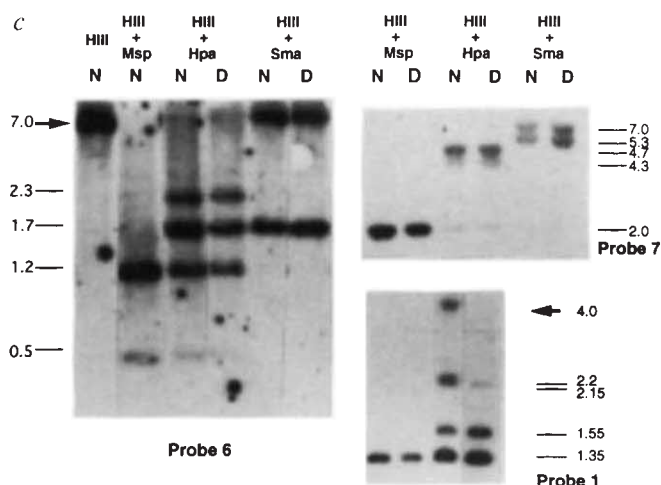
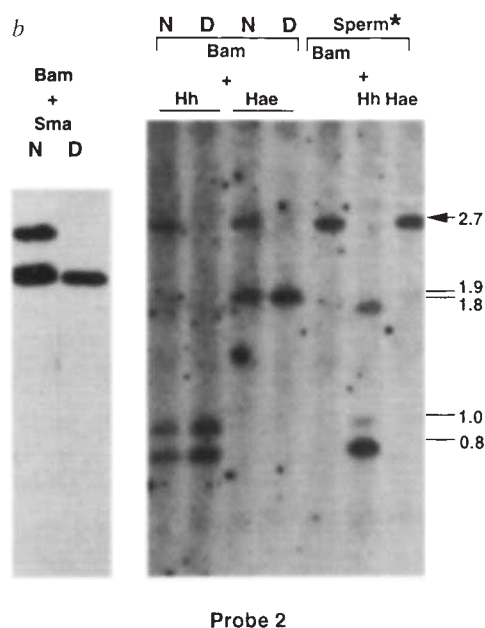
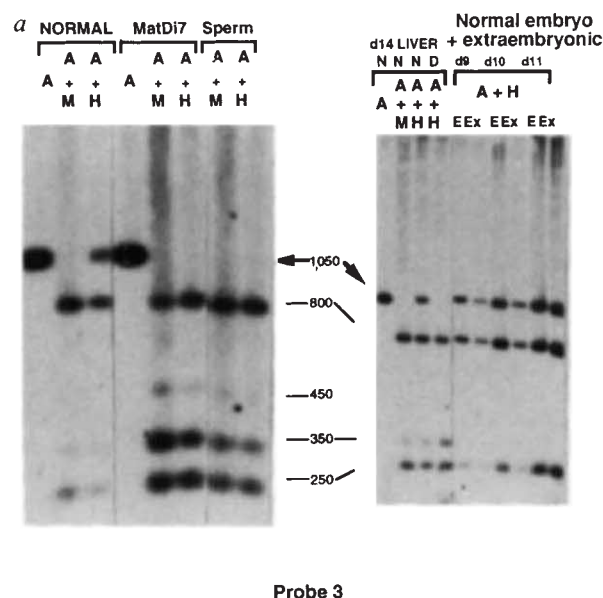


FIG. 3 Differential methylation of *H19* in MatDi7 and normal embryos. Southern blot analysis of MatDi7 (D), normal (N) and sperm DNAs, double-digested and hybridized with the probes numbered as shown in Fig. 2. **a**, Sites in and 5' to the promoter, showing digestion of both maternal chromosomes in MatDi7 (see Methods for restriction enzyme notation); in normal embryos one copy remains uncut. Sperm DNA and DNA from normal embryos (E) and extraembryonic tissue (Ex) at d9, d10 and d11 of gestation was also analysed. DNA from whole embryo and liver always gave the same results. **b**, Analysis of the two sites 5' to the promoter that show differential methylation, as well as the *SmaI* site in the promoter. The two sites are methylated in sperm DNA (see text). **c**, Sites in the 3' portion of the gene in the region of the enhancers show equivalent methylation of both parental chromosomes as do the two *HpaII* sites 5' to the promoter which are partially methylated in MatDi7 embryos.

METHODS. DNAs were double digested with *Apal* (A), *BamHI* (Bam) or *HindIII* (HIII) and the methylation sensitive enzymes, *MspI* (M), *HpaII* (H), *SmaI* (s), *HhaI* (Hh) or *HaeIII* (Hae) as indicated and hybridized with *H19* probes as shown. Autoradiographs **a**, **b** and **c** are representative blots and are reproducible for several different isolates. Southern blot analysis was carried out as described previously⁴. Molecular weights were determined using λ /HindIII- ϕ X174/*HaeIII* DNA digests as markers. Strains used in the analyses were as follows: MatDi7 and control littermates F_1 (T9Hc/+c $\times$ T9H/+c) (see Fig. 1 legend); sperm DNA isolated from adult epididymus was BALB/c or T9Hc/+c, and normal E and Ex were F_1 (CBA $\times$ C57Bl/6J).

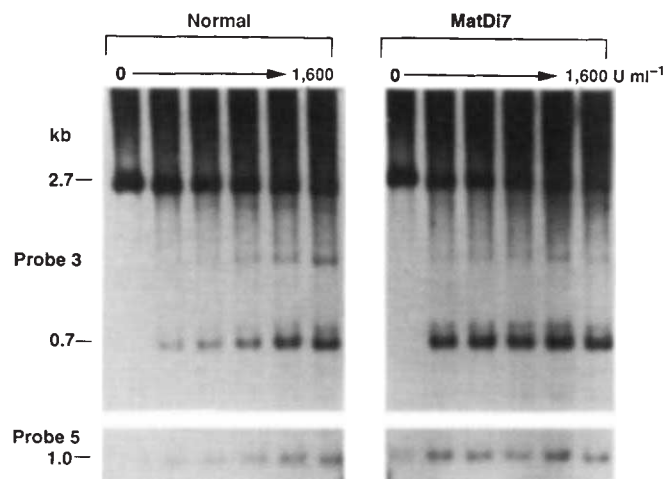


FIG. 4 The chromatin of MatDi7 embryos is more accessible to nucleases than that of normal embryos. Assay of *MspI* nuclease accessibility. Nuclei were isolated from d15 embryos and treated as described⁴ except that *MspI* (Stratagene) was added at 100, 200, 400, 800 and 1,600 units ml⁻¹. After termination of nuclease reactions, DNA was purified and digested with *BamHI*, subjected to Southern blot hybridization with probe 3 or control probe 5. Band intensities were analysed as described for Fig. 1.

90 kb upstream²⁹. Unlike *H19*, *Igf2* shows no detectable parental-origin-specific differences in methylation or chromatin compaction⁴. The proximity and overlapping temporal and spatial patterns of expression of these two genes may suggest that there is a functional and/or mechanistic relationship between them. If so, then a single region might be responsible for controlling the imprinting of the domain encompassing both *H19* and *Igf2*. It remains to be seen whether there is an association between the reciprocal imprinting of these two genes. □

Received 11 December 1992; accepted 29 March 1993.

- Poirier, F. *et al. Development* **113**, 1105–1114 (1991).
- Bartolomei, M. S., Zemel, S. & Tilghman, S. M. *Nature* **351**, 153–155 (1991).
- Ferguson-Smith, A. C., Cattanach, B. M., Barton, S. C., Beechey, C. V. & Surani, M. A. *Nature* **351**, 667–670 (1991).
- Sasaki, H. *et al. Genes Dev.* **6**, 1843–1856 (1992).
- Searle, A. G. & Beechey, C. V. *Genet. Res., Camb.* **56**, 237–244 (1990).
- DeChiara, T. M., Robertson, E. J. & Efstratiadis, A. *Cell* **64**, 849–859 (1991).
- Brunkow, M. E. & Tilghman, S. M. *Genes Dev.* **5**, 1092–1101 (1991).
- Bird, A. *Nature* **321**, 209–213 (1986).
- Cedar, H. *Cell* **53**, 3–4 (1988).
- Howlett, S. K. & Reik, W. *Development* **113**, 119–127 (1991).
- Kafri, T. *et al. Genes Dev.* **6**, 705–714 (1992).
- Li, E., Bestor, T. H. & Jaenisch, R. *Cell* **69**, 915–926 (1992).
- Monk, M. *Bioessays* **4**, 204–208 (1986).
- Grant, S. G. & Chapman, V. M. *A. Rev. Genet.* **22**, 199–233 (1988).
- Sasaki, H., Allen, N. D. & Surani, M. A. in *DNA Methylation: Molecular Biology and Biological Significance* (eds Jost, J. & Saluz, H.) 469–486 (Birkhauser, Basel, 1993).
- Lin, D. & Chinault, A. C. *Som. Cell molec. Genet.* **14**, 261–272 (1988).
- Antequera, F., MacLeod, D. & Bird, A. P. *Cell* **58**, 509–517 (1989).
- Pourcel, C., Tollais, P. & Farza, H. *J. Virol.* **64**, 931–935 (1990).
- Yoo-Warren, H., Pachnis, V., Ingram, R. S. & Tilghman, S. M. *Molec. cell. Biol.* **8**, 4707–4715 (1988).
- Surani, M. A. *et al. Development* suppl. 89–98 (1990).
- Antequera, F. & Bird, A. in *DNA Methylation: Molecular Biology and Biological Significance* (eds Jost, J. & Saluz, H.) 169–185 (Birkhauser, Basel, 1993).
- Monk, M. & Grant, M. *Development* suppl. 55–62 (1990).
- Boyes, J. & Bird, A. *Cell* **64**, 1123–1134 (1991).
- Levine, A., Cantoni, G. L. & Razin, A. *Proc. natn. Acad. Sci. U.S.A.* **89**, 10119–10123 (1992).
- Lewis, J. D. *et al. Cell* **69**, 905–914 (1992).
- Meehan, R. R., Lewis, J. D., McKay, S., Kleiner, E. L. & Bird, A. P. *Cell* **58**, 499–507 (1989).
- Comb, M. & Goodman, H. M. *Nucleic Acids Res.* **18**, 3975–3982 (1990).
- Holler, M., Westin, G., Jiriny, J. & Schaffner, W. *Genes Dev.* **2**, 1127–1135 (1988).
- Zemel, S., Bartolomei, M. S. & Tilghman, S. M. *Nature Genet.* **2**, 61–65 (1992).
- Heandler, B. & Hofer, E. *Eur. J. Biochem.* **190**, 477–482 (1990).
- Marzluff, W. F. & Huang, R. C. C. in *Transcription and Translation: A Practical Approach* (eds Hames, B. D. & Higgins, S. J.) 89–129 (IRL, Oxford, 1984).
- Gardiner-Garden, M. & Frommer, M. *J. molec. Biol.* **196**, 261–282 (1987).

ACKNOWLEDGEMENTS. We thank T. Koide, P. Rigby and C.-T. J. Chan for cDNA and cosmid clones; M. Norris, M. Wijgerde and C. Beechey for assistance; P. Jones and members of the M.A.S. and Reik Laboratories for discussion; and W. Reik for critically reading the manuscript. This work was supported by the AFRC and the Wellcome Trust.

Idiotypic/granulocyte-macrophage colony-stimulating factor fusion protein as a vaccine for B-cell lymphoma

Mi-Hua Tao* & Ronald Levy

Department of Medicine, Division of Oncology, School of Medicine, Stanford University, Stanford, California 94305, USA

* Present address: Institute of Biomedical Sciences, Academia Sinica, Taipei 11529, Taiwan, China

To produce a vaccine against cancer, antigens must be found that are preferentially expressed by tumour cells and can induce an immune response against the tumour. The variable regions of the immunoglobulin molecules expressed on malignant B cells (idiotypes) are tumour-specific, but are weak immunogens. To induce an immune response in animals or humans, the idiotype protein has therefore to be chemically coupled to a strongly immunogenic protein and mixed with an adjuvant^{1–8}. The resulting response can protect animals from subsequent tumour challenge, and cure animals with established tumours in combination with chemotherapy. Granulocyte-macrophage colony-stimulating factor (GM-CSF) augments antigen presentation in a variety of cells^{9–12}. Here we show that by fusing a tumour-derived idiotype to GM-

TABLE 1 Anti-idiotypic antibodies induced by immunization with different idiotype proteins

Immunogen	Dose (μg)	Anti-idiotypic antibody (μg ml ⁻¹ ± s.d.)	
		1st immunization	2nd immunization
Experiment 1			
Id	50	0	0
Id + rMoGM-CSF	50 ± 1.3	0	0
Id/MoGM	10	9.0 ± 7.0	25.4 ± 13.7
Id/HuGM	50	9.0 ± 5.0	26.3 ± 9.3
KLH-Id	50	2.7 ± 2.0	36.0 ± 23.3
Experiment 2			
Id	10	0	0
Id/HuGM	10	0	0
Id/MoGM	10	2.2 ± 1.8	60.7 ± 7.4
KLH-Id + adjuvant*	10	0.5 ± 0.7	51.2 ± 41.8

Groups of female C3H/HeN mice were immunized intraperitoneally with immunogens in phosphate-buffered saline unless otherwise indicated. Ten mice were included in each group except the group immunized with 50 μg Id/MoGM, which had five animals. Two weeks later, mice were boosted with the same antigens. KLH-Id was made by coupling KLH to Id using 0.1% glutaraldehyde⁴. The cytokine molarity of 1.3 μg recombinant murine GM-CSF (rMoGM-CSF) (gift from D. Rennick), is equivalent to 10 μg Id/MoGM. Serum was sampled by tail-bleeding 10 d after each immunization. The titres of antisera against native 38C13 idiotype (IgM, κ) were determined by ELISA using a cocktail of murine monoclonal antibodies (S3H5 (IgG1), S1C5 (IgG2a), S5A8 (IgG2b)¹⁴) as a standard. Data are expressed as the mean concentration ± standard deviation (s.d.).

* KLH-Id was administered with the adjuvant SAF-1 (ref. 29) in the first immunization and with incomplete SAF-1 (lacking the MDP component) in the second immunization.

CSF, it can be converted into a strong immunogen capable of inducing idiotype-specific antibodies without other carrier proteins or adjuvants and of protecting recipient animals from challenge with an otherwise lethal dose of tumour cells. This approach may be applicable to the design of vaccines for a variety of other diseases.

Genetic constructs for heavy and light chains were made as described in Fig. 1a. Plasmid p3079 was constructed with the coding sequence of heavy-chain variable region (VH38C) from the mouse B-cell tumour, 38C13 (ref. 13), inserted upstream of human IgG1 heavy-chain constant region gene. p3077 was created likewise to make the VL38C and human immunoglobulin κ chimaeric light chain. The p3159 and p3177 plasmids were constructed by generating a *Cla*I and a *Not*I site next to the last codon of the CH3 exon and inserting genetic fragments encoding murine GM-CSF and human GM-CSF, respectively. The heavy-chain vectors were then cotransfected with the light-chain vector by electroporation into an immunoglobulin non-secreting plasmacytoma Ag8.653 and selected for G418 resistance. Supernatants from resistant cells were screened for the presence of molecules expressing human IgG1, κ antigenic determinants by ELISA. The proteins made by p3079-, p3159- and p3177-transfected cells were designated as Id, Id/MoGM and Id/HuGM, respectively. All proteins were produced in tissue culture and purified by protein A chromatography.

These chimaeric tumour idiotype proteins were subjected to sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS–PAGE) (Fig. 1b). Under reducing conditions the protein bands of Id migrate with apparent M_r s of 58 and 27K, representing the heavy and light chains, respectively. Reduction of Id/MoGM and Id/HuGM also gave a 27K light chain, but both heavy-chain molecules migrated with an apparent M_r of 76K, indicating that they contained the GM-CSF tail. The heavy and light chains were assembled properly to give tetrameric proteins of M_r 160K (Id) and 219K (Id/MoGM and Id/HuGM), as can be seen in Fig. 1b on a non-reducing gel. The assembled fusion proteins were therefore dimeric with respect to GM-CSF.

The Id and the two GM-CSF fusion proteins all reacted with four different monoclonal anti-idiotypic antibodies raised against the native murine 38C13 idiotype protein¹⁴, indicating that the idiotype was intact (data not shown). To determine the functional activity of the murine GM-CSF portion of the fusion molecule, purified proteins were analysed for their ability to support the proliferation of a murine GM-CSF-dependent cell line NFS-60 (ref. 15). Id/MoGM clearly demonstrated the ability to stimulate the growth of NFS-60 cells in a dose-dependent manner (Fig. 2a). Moreover, the fusion protein exhibited a more potent cytokine activity than the recombinant murine GM-CSF even when corrected for the molarity of the cytokine. The presence of two GM-CSF molecules, each attached to one heavy chain, in Id/MoGM could increase its avidity to cellular receptors or, alternatively, could crosslink receptors¹⁶ and thus enhance its bioactivity. A similar enhancement in receptor affinity and biological activity were also reported for a fusion protein composed of GM-CSF and interleukin-3 (IL-3)¹⁷, whose receptors share a common β -chain^{18,19}. As expected, Id was completely negative in the GM-CSF bioassay, as was Id/HuGM, because human GM-CSF has no biological activity in murine cells²⁰. Figure 2b shows that the activity of Id/MoGM could be inhibited by a rat anti-murine GM-CSF monoclonal antibody but not by a control antibody, indicating that the cytokine activity is specific. Incubation of Id/MoGM and antibody against human γ -chain also arrested cell growth, indicating that GM-CSF activity is in the form of a fusion molecule. These results clearly demonstrate that the GM-CSF moiety of Id/MoGM was in a fully functional configuration.

Mice immunized with Id and its GM-CSF-derived fusion proteins were analysed for their anti-idiotypic antibody responses. Id did not induce detectable anti-idiotypic antibodies (Table 1). The carrier or helper effects of human immunoglobulin-constant regions were therefore negligible. Conversely, Id/MoGM induced significant amounts of anti-idiotypic antibodies in all immunized animals. Animals immunized with a simple mixture of Id and recombinant murine GM-CSF produced no anti-idiotypic response, indicating that fusion of GM-CSF to tumour idiotype was required for enhanced immunogenicity. The fact that no anti-idiotypic antibodies were induced by the similar but biologically inert human GM-CSF fusion protein indicates that the murine GM-CSF on Id/MoGM required cytokine activity for its immune-enhancing function and was not simply acting as a carrier protein. Compared with mice immunized with Id chemically conjugated to the strong carrier protein keyhole limpet haemocyanin (KLH), with or without adjuvants, titres of anti-idiotypic antibodies induced by Id/MoGM were higher in the early response and about equal in the later response. We found that Id/MoGM worked very efficiently in eliciting antibody responses; as little as 0.05 μ g could induce significant titres of anti-idiotypic antibodies (data not shown). The immune responses are specific because no reactivity occurred against other proteins such as KLH or 4C5 (an irrelevant mouse IgM) (data not shown). Furthermore, in fluorescent-activated cell sorting (FACS) analysis we could show that the immune sera from mice vaccinated with Id/MoGM specifically recognized 38C13 tumour cells but not an immunoglobulin-negative tumour variant^{21,22} (data not shown).

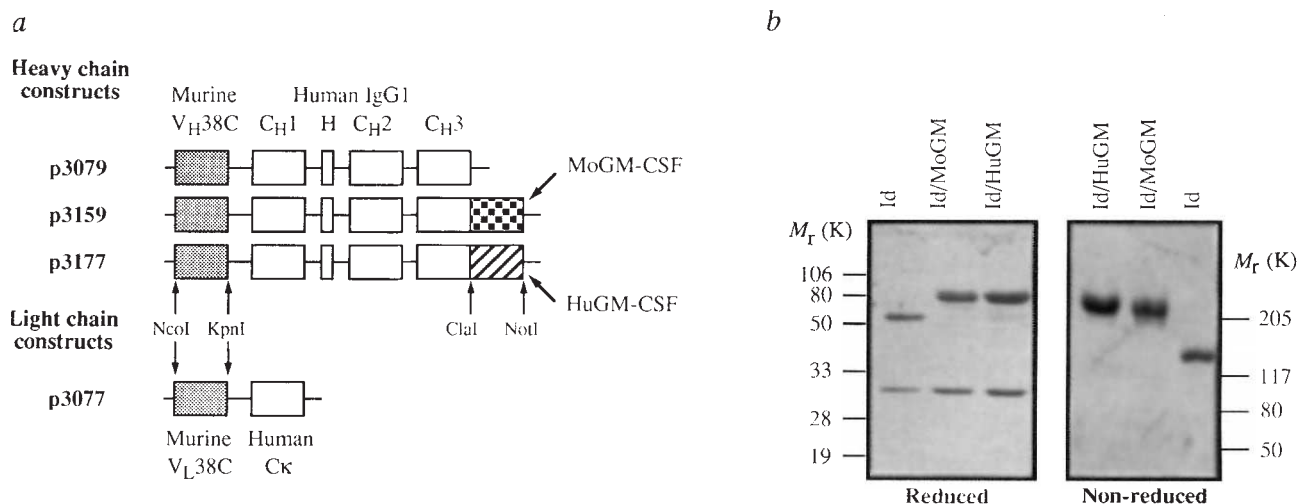


FIG. 1 Construction of tumour idiotype protein and its GM-CSF derivatives. *a*, Schematic representation of the heavy-chain (p3079, p3159, p3177) and light-chain (p3077) genetic constructs. The coding sequences for 38C13 heavy-chain variable region (V_H38C) and light-chain variable region (V_L38C) were ligated upstream of the human γ 1 heavy-chain constant region C_H1 exon and κ light-chain constant region exon (C_κ), respectively. Exons depicted as boxes. The genetic fragment encoding the mature peptide of murine GM-CSF (MoGM) or human GM-CSF (HuGM) was ligated 3' to the C_H3 exon in p3159 and p3177. *b*, Coomassie brilliant blue-stained SDS-polyacrylamide gel of purified Id, Id/MoGM, and Id/HuGM under reducing and non-reducing conditions.

METHODS. The *Eco*RI segment which contains the V_H gene of the immunoglobulin heavy-chain expression vector, pSV2ΔHgtVHDNSHuG1, and the *Hind*III segment which contains the V_L gene of the light-chain vector, pSV184ΔHneoVLDNSHuK³⁰, were cloned into pUC9. Site-directed mutagenesis was used to introduce *Nco*I and *Kpn*I sites in the leader and joint segments to produce plasmids p3071 and p3074. V_H38C and V_L38C were obtained by reverse transcription and polymerase chain reaction (PCR) amplification of RNA derived from 38C13 tumour cells and replaced the corresponding VHDNS and VLDNS in p3071 and p3074 to produce p3075 and p3076. The V_H38C-containing *Eco*RI fragment and the V_L38C-containing *Hind*III fragment were then inserted into the respective expression vectors to produce p3079 and p3077. To construct the immunoglobulin heavy

chain/GM-CSF chimaeric molecules, site-directed mutagenesis was used to generate a *Cla*I site between the last codon and the stop codon of the C_H3 exon and a *Not*I site 3' to the stop codon in p3079 to produce plasmid p3143. The mature murine GM-CSF sequence was obtained by reverse transcription and PCR amplification of RNA derived from about 1×10^8 mouse spleen and lymph node cells stimulated for 24 h with 5 ng ml⁻¹ of phorbol myristate acetate (Sigma) and 5 μ g ml⁻¹ concanavalin A (Sigma). The fragment containing mature human GM-CSF sequence was generated by PCR from plasmid p91023 obtained from American Type Culture Collection (ATCC). The upstream primers contained two Gly codons 3' to the *Cla*I site, and the downstream primers contained a stop codon 5' to the *Not*I site. *Cla*I-*Not*I fragments of the PCR products were gel-purified and inserted between *Cla*I and *Not*I sites of p3143 to produce p3159 and p3177. Plasmids p3079, p3159 and p3177 were cotransfected with p3077 into P3X63Ag8.653 plasmacytoma cell line by electroporation and screened for immunoglobulin-producing clones following previously described procedures³⁰. The highest producers of each construct were subcloned, expanded in 3-litre tissue culture flasks (Baxter) at 37 °C, 5% CO₂ for 10 days, and purified by protein A chromatography. Protein was eluted from the column with 0.1 M glycine, 0.15 M NaCl, pH 2.4, buffer and brought to neutral pH with 0.5 M sodium phosphate, pH 8. Purified proteins were electrophoresed on a 12.5% (reduced) and a 5% (non-reduced) SDS-polyacrylamide gels and stained with Coomassie brilliant blue.

Two weeks after the last immunization, mice were challenged with 38C13 tumour cells; the results are presented in Fig. 3. In experiment 1 (Fig. 3a), mice immunized with Id or with a mixture of Id and recombinant murine GM-CSF did not show any protection; all of these mice died between day 16 and day 23. In contrast, mice immunized with 10 μ g or 50 μ g Id/MoGM fusion protein had greatly increased survival time and yielded 40% and 60% long-term survivors, respectively ($P < 0.001$ and $P < 0.005$). These mice were better protected than those injected

with KLH-conjugated Id, which resulted in 30% long-term survivors ($P < 0.001$). In a separate tumour-challenge experiment (Fig. 3b), we observed similar tumour immunity induced by Id/MoGM, whereas the same amount of Id/HuGM fusion protein induced no protection. We showed that these long-term survivors (>120 days after tumour challenge) were free of residual or dormant tumour cells by several criteria, including FACS analysis, *in vitro* culture and *in vivo* transfer of splenocytes (data not shown). The tumours that did grow in the immunized animals were examined for surface idiotype expression and were found to be positive.

Others have reported that administration of free GM-CSF increased antibody responses in mice against sheep red blood cells under certain conditions⁹. We also found that coimmunization of the Id/MoGM with KLH or hen egg lysozyme, both of which are immunogenic in mice, resulted in a slight enhancement of antibody titre to the unrelated antigens (data not shown). But to a weak immunogen, physical linkage to GM-CSF is

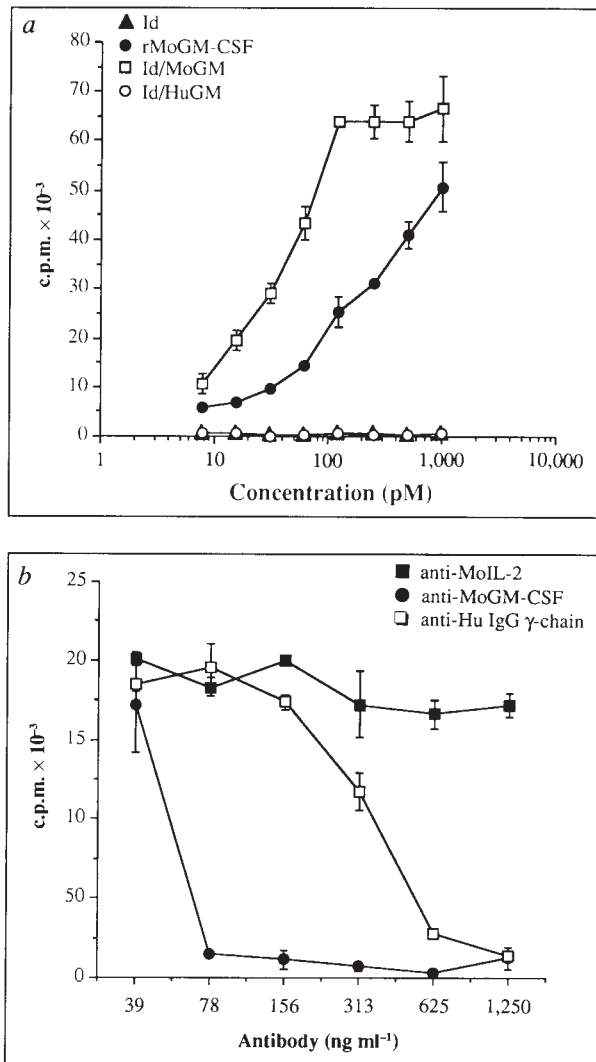


FIG. 2 Assay of the biological activity of the Id/GM-CSF fusion proteins. *a*, Id, Id/MoGM, Id/HuGM and recombinant murine GM-CSF were assayed for the ability to stimulate the proliferation of NFS-60 cells. *b*, Effect of anti-human γ -chain antibody, anti-murine GM-CSF antibody, and anti-murine IL-2 antibody on the growth of NFS-60 cells in medium containing 200 pM of Id/MoGM. Cell proliferation was measured by [³H]thymidine incorporation. All determinations were made in triplicate. The standard deviation of each determination is presented on the graph.

METHODS. In the proliferation assay, murine GM-CSF-dependent NFS-60 cells¹⁵ were washed three times and plated at 5×10^3 cells per well in 0.2 ml Iscove modified medium containing 10% heat inactivated fetal bovine serum and the indicated amounts of test samples. Recombinant murine GM-CSF was purchased from Genzyme. In the inhibition assay, cells were cultured with 200 pM of Id/MoGM and various amounts of either goat anti-human γ -chain antibody (TAGO), rat anti-murine GM-CSF antibody (Genzyme), or rat anti-murine IL-2 antibody (Genzyme). Cells were incubated for 18 h, then 1 μ Ci [³H]thymidine (Amersham) was added to each well and incubation continued for an additional 6 h. Cells were then collected and the amount of radioactivity determined in a liquid scintillation counter.

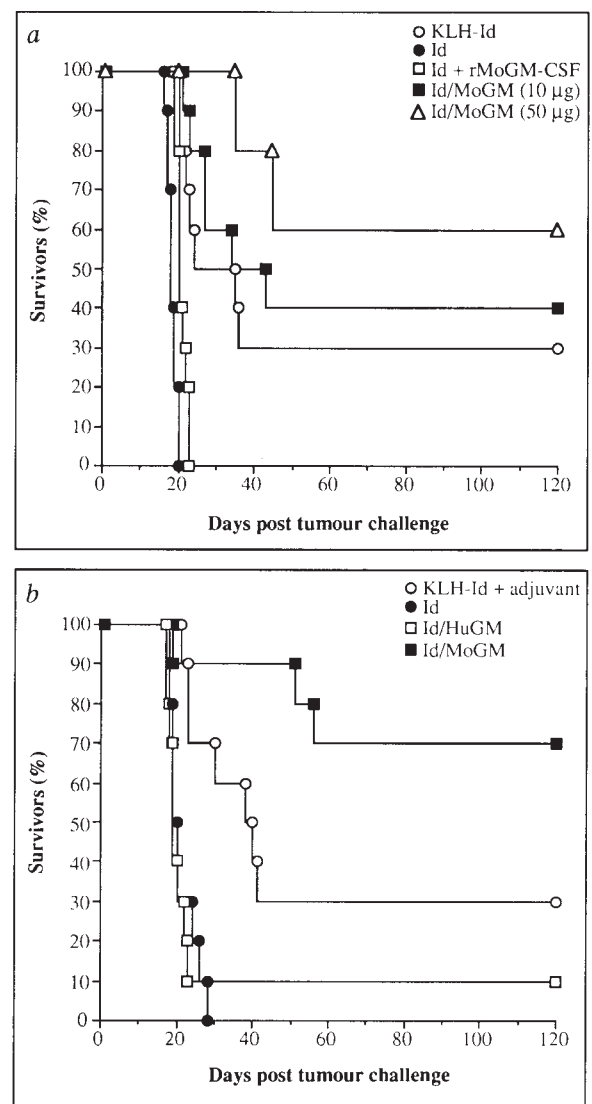


FIG. 3 Tumour immunity induced by immunization with Id and its GM-CSF derived fusion proteins. Mice immunized as described in Table 1 legend were challenged with 38C13 tumour cells and the percentage of survivors recorded. *a*, Experiment 1; *b*, experiment 2.

METHODS. 38C13 tumour cells were maintained in culture in RPMI 1640 medium containing 10% fetal calf serum and 5×10^{-5} M 2-mercaptoethanol. Two weeks after the second immunization, mice were challenged with 200 38C13 tumour cells, i.p. The survival of challenged mice was followed and the results analysed for significance by Wilcoxon's two-sample test.

required to get optimal immune responses. Although Id/MoGM can elicit strong immunity against 38C13 idiotype, a simple mixture of Id/MoGM with another idiotypic protein 4C5 induces no 4C5-specific antibodies. It has been shown recently that local production of cytokines by genetically engineered tumour cells decreased the tumorigenicity and elicited protective immune responses against the parental tumours²³⁻²⁷. These results highlight the importance of local cytokine effects on immune responses. Compared to the cytokine gene-targeted tumour method, our approach of using tumour antigen/GM-

CSF fusion proteins as vaccines has the advantage of using a chemically defined entity as the vaccine. A fusion protein of interleukin-2 and herpes simplex virus type 1 (HSV-1) glycoprotein D was also reported to enhance both humoral and cellular immunity against HSV-1 (ref. 28). These results indicate that the bioactivity of cytokines can obviate the requirement for potentially toxic immunological adjuvants. Such antigen-cytokine fusion proteins may be an improvement over current approaches to tumour vaccines as well as vaccines against other diseases. □

Received 3 December 1992; accepted 29 March 1993.

1. Lynch, R. G., Graff, R. J., Sirisinha, S., Simms, E. S. & Eisen, H. N. *Proc. natn. Acad. Sci. U.S.A.* **69**, 1540-1544 (1972).
2. Sugai, S., Palmer, D. W., Talal, N. & Witz, I. P. *J. exp. Med.* **140**, 1547-1558 (1974).
3. Freedman, P. M., Autry, J. R., Tokuda, S. & Williams, R. J. *J. natn. Cancer Inst.* **56**, 735-740 (1976).
4. Campbell, M. J. *et al. J. Immun.* **139**, 2825-2833 (1987).
5. Kaminski, M. S., Kitamura, K., Maloney, D. G. & Levy, R. *J. Immun.* **138**, 1289-1296 (1987).
6. George, A. J., Folkard, S. G., Hamblin, T. J. & Stevenson, F. K. *J. Immun.* **141**, 2168-2174 (1988).
7. Campbell, M. J., Esserman, L. & Levy, R. *J. Immun.* **141**, 3227-3233 (1988).
8. Kwak, L. W. *et al. New Engl. J. Med.* **327**, 1209-1215 (1992).
9. Morrissey, P. J., Bressler, L., Park, L. S., Alpert, A. & Gillis, S. *J. Immun.* **139**, 1113-1119 (1987).
10. Heuffer, C., Koch, F. & Schuler, G. *J. exp. Med.* **167**, 700-705 (1988).
11. Fischer, H. G., Frosch, S., Reske, K. & Reske, K. A. *J. Immun.* **141**, 3882-3888 (1988).
12. Smith, P. D., Lamerson, C. L., Wong, H. L., Wahl, L. M. & Wahl, S. M. *J. Immun.* **144**, 3829-3834 (1990).
13. Bergman, Y. & Haimovich, J. *Eur. J. Immun.* **7**, 413-417 (1977).
14. Maloney, D. G., Kaminski, M. D., Burowski, D., Haimovich, J. & Levy, R. *Hybridoma* **4**, 191-209 (1985).
15. Holmes, K. L., Palaszynski, E., Fredrickson, T. N., Morse, H. C. & Ihle, J. N. *Proc. natn. Acad. U.S.A.* **82**, 6687-6691 (1985).
16. Ulrich, A. & Schlessinger, J. *Cell* **61**, 203-212 (1990).

17. Curtis, B. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 5809-5813 (1991).
18. Nicola, N. A. & Metcalf, D. *Cell* **67**, 1-4 (1991).
19. Lopez, A. F., Elliott, M. J., Woodcock, J. & Vadas, M. A. *Immun. Today* **13**, 495-500 (1992).
20. Luzzati, A. J., Quon, D. H. & Golde, D. W. *Blood* **57**, 13-21 (1981).
21. Carroll, W. L., Starnes, C. O., Levy, R. & Levy, S. *J. exp. Med.* **168**, 1607-20 (1988).
22. Starnes, C. O. *et al. J. Immun.* **141**, 333-339 (1988).
23. Tepper, R. I., Pattengale, P. K. & Leder, P. *Cell* **57**, 503-512 (1989).
24. Fearon, E. R. *et al. Cell* **60**, 397-403 (1990).
25. Gansbacher, B. *et al. Cancer Res.* **50**, 7820-7825 (1990).
26. Blankenstein, T. *et al. J. exp. Med.* **173**, 1047-1052 (1991).
27. Hock, H., Dorsch, M., Diamantstein, T. & Blankenstein, T. *J. exp. Med.* **174**, 1291-1298 (1991).
28. Hinuma, S., Hazama, M., Mayumi, A. & Fujisawa, Y. *FEBS Lett.* **288**, 138-142 (1991).
29. Allison, A. C. & Byars, N. E. *J. Immun. Meth.* **95**, 157-168 (1986).
30. Shin, S. U. & Morrison, S. L. *Meth. Enzym.* **178**, 459-476 (1989).

ACKNOWLEDGEMENTS. We thank D. Rennick (DNAX) for NFS-60 cells and recombinant murine GM-CSF; S. L. Morrison (University of California, Los Angeles) for plasmids pSV2ΔHspVHDNSHuG1 and pSV184ΔHneoVLDNSHuX; and S. Levy for many helpful discussions. M.H.T. is a recipient of The Jane Coffin Childs Memorial Fund postdoctoral fellowship. R.L. is a Clinical Research Professor of the American Cancer Society. This work was supported by grants from the NIH-USPHS.

Virus persistence in acutely infected immunocompetent mice by exhaustion of antiviral cytotoxic effector T cells

Demetrius Moskophidis, Franziska Lechner, Hanspeter Pircher & Rolf M. Zinkernagel

Institute for Experimental Immunology, Department of Pathology, University of Zurich, Sternwartstrasse 2, CH-8091 Zurich, Switzerland

VIRUSES that are non- or poorly cytopathic have developed various strategies to avoid elimination by the immune system and to persist in the host¹⁻³. Acute infection of adult mice with the noncytopathic lymphocytic choriomeningitis virus (LCMV) normally induces a protective cytotoxic T-cell response that also causes immunopathology⁴⁻⁷. But some LCMV strains (such as DOCILE⁸ (LCMV-D) or CI-13 Armstrong (CI-13)⁹) derived from virus carrier mice tend to persist after acute infection of adult mice without causing lethal immunopathological disease^{4,5}. Tendency to persist correlates with tropism^{8,10}, rapidity of virus spread⁸ and virus mutations^{11,12}. We report here that these LCMV isolates may persist because they induce most of the specific antiviral CD8⁺ cytotoxic T cells so completely that they all disappear within a few days and therefore neither eliminate the virus nor cause lethal immunopathology. The results illustrate that partially and sequentially induced (protective) immunity or complete exhaustion of T-cell immunity (high zone tolerance) are quantitatively different points on the scale of immunity; some viruses exploit the latter possibility to persist in an immunocompetent host.

Because transplacentally or neonatally LCMV-infected mice become virus carriers and lack LCMV-specific cytotoxic T cells (CTLs)¹³⁻¹⁵, we evaluated the fate of anti-LCMV CTL responses after acute infection of adult C57BL/6 mice leading to LCMV-persistence. Mice infected with 10² plaque forming units (PFU) of LCMV-D cleared virus completely by day 15 to 30 in spleens and thymus (Fig. 1a, b); they produced conventional CTL

responses (Fig. 1c) and *in vitro* restimulatable CTL precursors (CTLp) (Fig. 1d) peaking around day 9, and declining to memory levels by day 15 to 40 (ref. 16). In contrast, 10⁷ PFU of LCMV-D induced only a weak primary CTL response measurable on day 6 and *in vitro* restimulatable CTLp disappeared after day 12 (Fig. 1). Virus persisted in all organs tested up to 280 days. Inoculation of 10⁴ PFU of LCMV-D is a critical level of infection because about half of the mice failed to clear the virus and exhibited no CTLp after day 40-120. Thus early and complete induction of anti-viral CTLs correlated with the disappearance of CTL activity and with LCMV persistence.

LCMV-D titres, CTL and CTLp kinetics were comparable in adult sham-operated and in adult thymectomized mice (data not shown); thus establishment of LCMV-D persistence in adult mice depended on peripheral not central exhaustion/tolerance of T cells. But in addition the results with euthymic mice imply that LCMV in the thymus (Fig. 1) further maintained established tolerance^{13,17} by clonal deletion of newly generated LCMV-specific CTL. Exhaustion of anti-LCMV CTLp was specific because after double infection all CTLp against LCMV were lost but CTLp frequencies against vaccinia virus were preserved (Table 1).

The disappearance of LCMV-specific CTL activity after an overwhelming LCMV-D infection may reflect anergy or peripheral clonal deletion of LCMV-specific CTL. Therefore, the fate of these CTL was evaluated by adoptive transfer of donor T cells from mice expressing a transgenic LCMV-specific T-cell receptor (TCR)¹³; these indicator CTL were then analysed in recipients directly by staining and flow cytometrical analysis (Fig. 2Aa, Ba, Ca), and functionally by testing their CTL activities after restimulation *in vitro* (Fig. 2Ab, Bb, Cb) and by determining LCMV-D titres in spleen and thymus (Fig. 2Ac, Bc, Cc). Transfer of 10³ transgenic TCR⁺ T cells to recipient C57BL/6 together with 10⁷ PFU of LCMV-D led to a vigorous expansion of transgenic T cells up to day 6 followed by a rapid decline to undetectable levels by day 15 (Fig. 2Aa). The kinetics of endogenous and of transgenic TCR⁺ restimulatable CTLp were indistinguishable but comparable to that of infected non-transfused controls (compare Fig. 2Ab with Aa, and data not shown).

TABLE 1 High-dose LCMV-Docile infection of C57BL/6 mice exhausts anti-LCMV CTLs specifically

Responder spleen cells from mice infected 20 days previously with:	Stimulator cells infected with:	CTL-activity (LU per spleen, $\times 100$) tested on MC57 (H-2 ^b)-target cells infected with:		
		LCMV-D	Vaccinia virus	Uninfected
LCMV-D (10^7 PFU)	LCMV-D	<0.2	ND	<0.2
LCMV-D (10^7 PFU) Vaccinia virus (2×10^6 PFU)	LCMV-D	<0.2	ND	<0.2
LCMV-D control (10^2 PFU)	LCMV-D	$2,285 \pm 38$	ND	<0.2
Vaccinia virus (2×10^6 PFU)	Vaccinia virus	ND	307 ± 4	3 ± 1
LCMV-D (10^7 PFU) + Vaccinia virus (2×10^6 PFU)	Vaccinia virus	ND	256 ± 29	1 ± 0.7

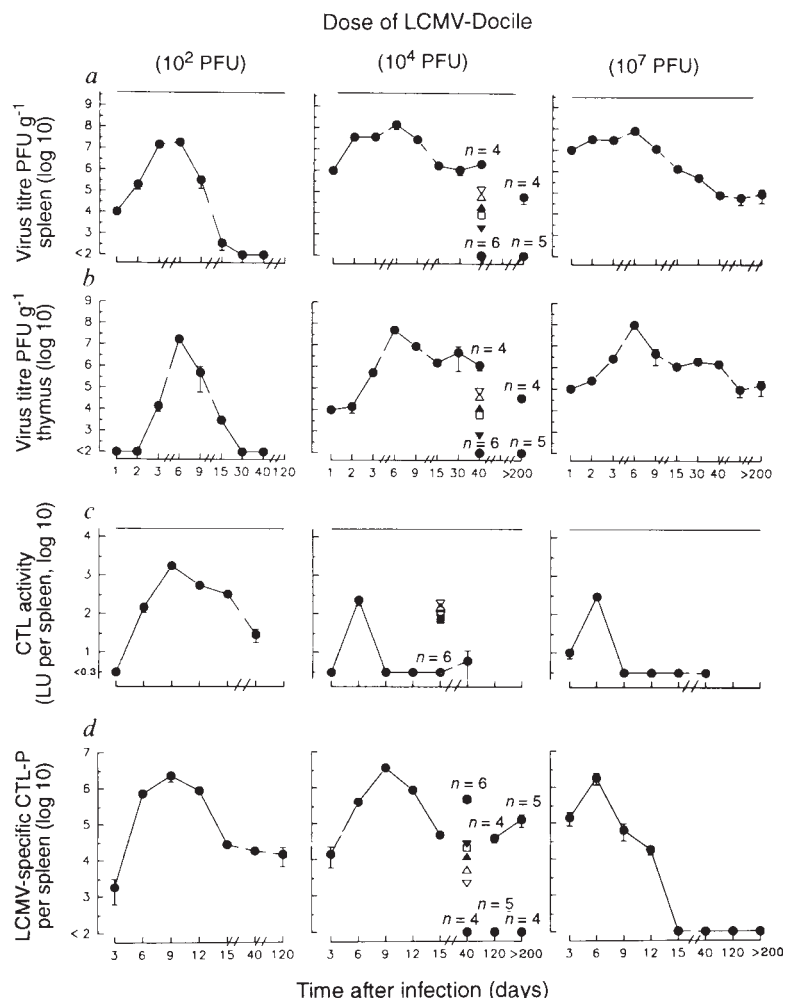
Mice were simultaneously infected with LCMV-D (10^7 PFU intravenously) and the poorly replicating vaccinia virus (2×10^6 PFU intravenously). Control mice were infected with LCMV-D (10^7 PFU) alone or with vaccinia virus (2×10^6 PFU) alone. Twenty days later the restimulatable antiviral CTL activities were evaluated *in vitro*. Various numbers of spleen cells (4×10^6 – 1.25×10^5 cells per well) were cultivated for 5 days in 24-well plates with irradiated, LCMV-D or vaccinia virus-infected macrophages as stimulators (5×10^5 cells per well) in medium rich in interleukins. The CTL activity was detected in a 4–5 h ^{51}Cr release assay on MC57G(H-2^b) target cells infected with LCMV-D or vaccinia virus or uninfected. Cytotoxic activity was expressed as lytic units (LU) per spleen on the basis of 3–5 mice $\pm$ s.e.m. per group.

If transferred on day 15 after infection with 10^7 PFU of LCMV-D, at a time when endogenous CTLs and CTLp had already disappeared, directly stainable transgenic TCR⁺ T cells (Fig. 2Ba) and their restimulatable specific cytolytic activity (Fig. 2Bb), increased for a few days and then completely disappeared by about day 15 after transfer. This was not due to clearance of LCMV because wild-type virus persisted at comparable titres in all mice (Fig. 2Ac, Bc). Transgenic TCR⁺ T cells from recipients on day 3 or 6 after transfer could be restimulated *in vitro* but not on day 10 or 15 after transfer (Fig. 2Ab, Bb). This was surprising for the day 10 T cells because they exhibited

normal staining levels of TCR (Fig. 2Aa, Ba); thus they could be called 'anergic'. Because the density of CD3, CD8 or TCR on transgenic T cells did not change during the course of infection (Fig. 3), downregulation of TCR or of coreceptors^{18,19} does not explain the results. When TCR⁺ transgenic T cells were cotransferred to C57BL/6 recipients infected with a low dose of LCMV-D (10^2 PFU), transgenic and endogenous CTL expanded enormously from an estimated 100 transgenic CTL⁺ T cells per spleen to 4×10^7 on day 9 and then dropped to about 7×10^5 by day 15 (Fig. 2Ca); restimulatable CTLp persisted (Fig. 2Cb) and LCMV-D was cleared (Fig. 2Cc). Thus peripheral clonal

FIG. 1 Exhaustion of LCMV-specific CTLs by widely replicating virus in immunocompetent mice. Kinetics of LCMV-D titres in the spleen (a) or thymus (b) of C57BL/6J infected intravenously with the indicated doses of 10^2 , 10^4 , 10^7 PFU LCMV-D. In parallel, primary CTL responses *ex vivo* (c) or numbers of restimulatable CTLp per spleen (d) were assessed. The different symbols used on day 40 in the middle column (10^4 PFU) represent values of individual mice (○, ▽, △, □, ■, ▲, ▼) or groups of 4–6 mice (●); all other ● values are means of 5–10 mice $\pm$ s.e.m.

METHODS. LCMV-Docile (a variant isolated from an LCMV-WE (UBC)-carrier mouse) was originally obtained from C. J. Pfau, Troy, NY⁸. Titres are PFU g⁻¹ organ, mean of 5–10 mice $\pm$ s.e.m. The primary antiviral CTL activity in the spleen of infected mice was tested directly in a 4–5 h ^{51}Cr release assay at an effector:target ratio of 70:1, 27:1, 9:1 and 3:1, and the values were expressed as standard lytic units (LU) per spleen on the basis of 5–10 mice $\pm$ s.e.m. per group⁷. In parallel and in the same mice the antiviral precursor activity was determined under limiting dilution conditions as has been described in detail previously¹⁶. The CTLp activity was expressed as total numbers of CTLp per spleen.



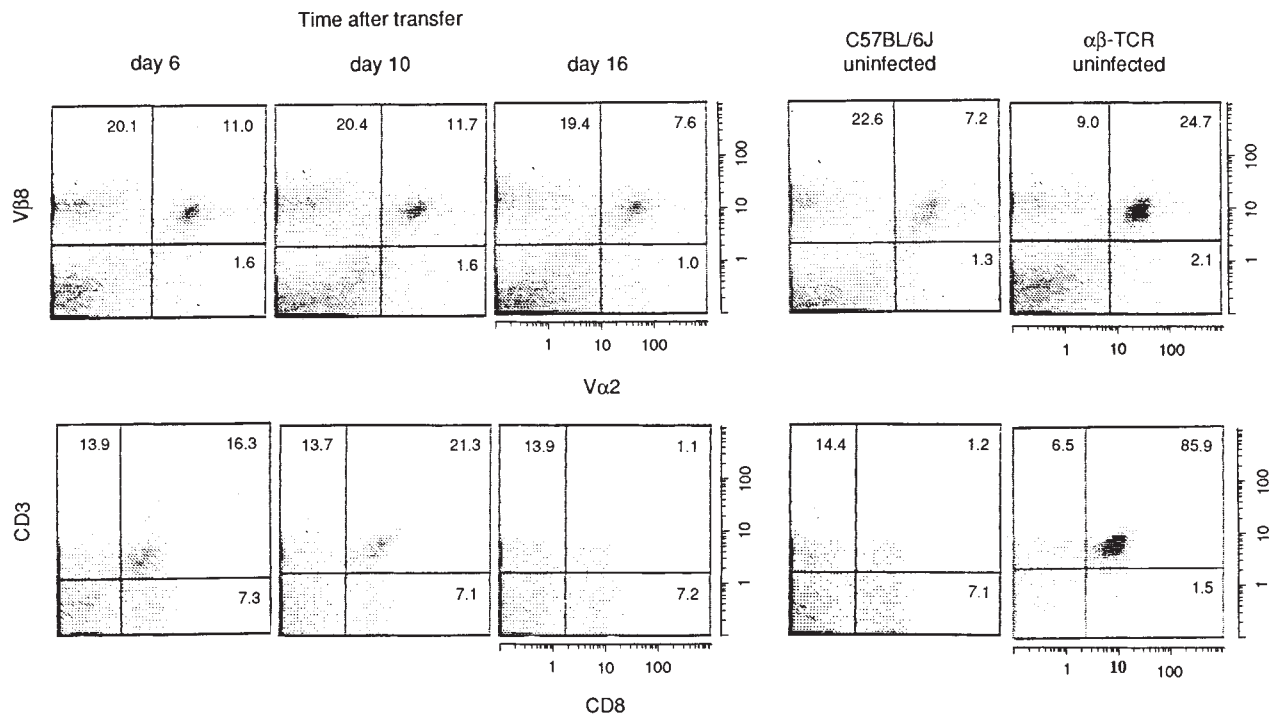
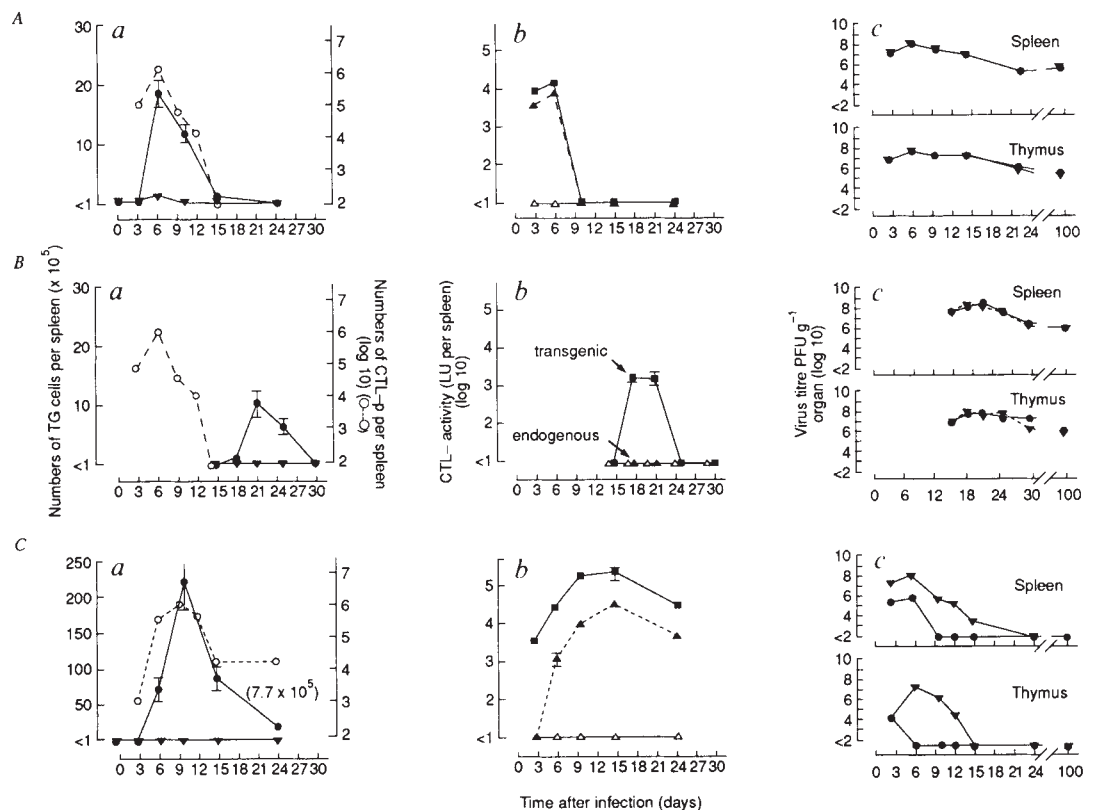


FIG. 2 Deletion of virus-specific CTLs in acutely LCMV-D-infected adult mice that become persistently infected. An optimal number of 10^3 transgenic TCR⁺ T cells (TG) specific for LCMV-GP amino acids 32–42 and H-2D^b were transferred together with 10^7 PFU LCMV-D into C57BL/6J-recipient mice to compare kinetics of endogenous and transferred transgenic CTLs (A). Alternatively cells were transferred on day 15 after infection (B), at a time when the endogenous CTL response had already functionally disappeared and adoptively transferred transgenic CTLs could therefore be specifically monitored. A third group of recipients of 10^3 transgenic TCR⁺ T cells was simultaneously infected with a low dose of LCMV-D (10^2 PFU) (C). Transgenic TCR⁺ CD8⁺ T cells were detected directly (before restimulation) by flow cytometric analysis using TCR-specific monoclonal antibodies (triple staining for Vα2, Vβ8 and CD8, for detail see legend to Fig. 3). Symbols in panels Aa, Ba, Ca represent absolute numbers of stainable transgenic TCR⁺ cells in infected recipient mice (●) or in nontransfused control C57BL/6J mice infected with 10^7 PFU of LCMV-D (▼). The functional status of endogenous CTLs (○) was determined in a separate group of C57BL/6J mice infected with 10^7 PFU (Aa, Ba) or 10^2 PFU (Ca) of LCMV-D alone. The function of transferred cells was evaluated in the same spleen cells used for direct TCR staining by determining their antiviral CTL capacity after they had been restimulated for 5 days (for methods see Table 1) (Ab, Bb, Cb). The CTL



activity was determined on MC57-G (H-2^b) target cells, uninfected (Δ), LCMV-D-infected ($\blacksquare$) or LCMV 8.7-infected ($\blacktriangle$). LCMV 8.7 virus exhibits a mutation in the transgenic TCR-specific epitope (amino-acids 32–42 of LCMV-GP), therefore 8.7-infected H-2^b target cells are not lysed by TCR-transgenic effector T cells⁴¹. Virus titres were determined in the spleens and thymus of the same LCMV-D-infected recipients of transgenic TCR⁺ T cells (●) or of LCMV-D-infected controls (▼) (Ac, Bc, Cc). Data points are means $\pm$ s.e.m. of 3–5 mice.

FIG. 3 Surface phenotype of transferred transgenic TCR⁺ CTL in persistently infected mice. 10³ transgenic TCR⁺ T cells were transferred on day 15 after infection with 10⁷ PFU LCMV-D into C57BL/6 mice (Fig. 2B) and their kinetics were followed after direct staining by flow cytometry. Top, expression of the transgenic TCR (V α 2/V β 8) on CD8⁺ splenocytes at the indicated time points after transfer. Splenocytes from normal C57BL/6 mice and TCR-transgenic mice were used as controls. Bottom, CD3 and CD8 expression of the same samples.

METHODS. Transgenic TCR expression was determined by triple staining: cells were first incubated with unconjugated anti-V α 2 monoclonal antibody⁴² followed by phycoerythrin (PE)-conjugated goat anti-rat immunoglobulin (Southern Biotechnology, Birmingham). After a blocking step with normal rat immunoglobulin, the staining was completed with anti-CD8-FITC (Becton Dickinson) and biotinylated anti-V β 8 (Pharmingen) and revealed by avidin-RED 613 (Gibco, BRL). CD3/CD8 double staining was done with unconjugated anti-CD3 monoclonal antibody⁴³ and FITC-conjugated goat anti-rat immunoglobulin (CALTAG, San Francisco). After a blocking step, it was completed by biotinylated anti-CD8 (Becton-Dickinson) and finally by avidin-PE (TAGO, Burlingame). Viable cells were gated by a combination of forward light scatter and 90° side scatter and were analysed on an EPICS profile analyser (Coulter, Hialeah, FL) with a scale extending over 4 log₁₀.

deletion of CTL was dependent on virus dose and kinetics. Additional experiments showed that exhaustion depended on the LCMV isolate (LCMV-D or CI-13 deleted CTL, but LCMV-Armstrong did not) and on major histocompatibility complex (MHC) and non-MHC genes of the host (not shown).

This drastic example of peripheral deletion of a well defined class I restricted antigen-specific effector of T-cell response adds to examples of anergy^{20,21}, partial disappearance of T cells^{19,22} or downregulation of either TCR and/or of coreceptor^{18,19} expression. Note that in contrast to these results, TCR-expressing T cells specific for superantigens²¹⁻²⁴ or for the male H-Y antigen¹⁹ have been found in frequencies well above background after day 15.

Why do exhaustively induced CTLs disappear? Because transferred transgenic CTLs expanded vigorously, even at times when the endogenous response had already disappeared (on day 15, Fig. 2B) interleukins^{25,26} or antigen presentation^{27,28} apparently do not seem to be the limiting factor. Also, attempts to prevent exhaustion by repeated infusions of interleukin-1, interleukin-2, interferon- γ or concentrated interleukin-rich supernatants have been unsuccessful (data not shown). The possibility that specific CTLs lysed CTLs because they were infected by LCMV²⁹ seems unlikely because of the demonstrated specificity (Table 1) and because neither endogenous nor transgenic CTLs were found here (or earlier³⁰) to be infected by LCMV (data not shown).

Together, these data show that, depending on the virus isolate and dose, most or all antiviral CTLs may become induced if noncytopathic viruses spread early, rapidly and widely in the host; then CTLs proliferate and disappear completely after a short period of 3-5 days of 'anergy'. The data suggest that T-cell memory represents continuous induction of T cells by antigen depots. Our findings explain mechanistically, but not yet molecularly, the puzzle of why high-dose infection with LCMV causes low CTL responses^{8,9,31-33} and after intracerebral infection (or after transient immunosuppression^{34,35}) fails to cause lethal choriomeningitis in adult mice^{4,5}. A similar mechanism of exhaustion of T-cell responses against widely spreading noncytopathic viruses may lead to a hepatitis B virus carrier status in a few immunocompetent men³⁶, or the disappearance of anti-HIV CTLs during final AIDS^{37,38}, may induce the veto phenomenon³⁹ or cause high zone tolerance to model antigens⁴⁰. □

6. Cole, G. A., Nathanson, N. & Prendergast, R. A. *Nature* **238**, 335-337 (1972).
7. Zinkernagel, R. M. & Doherty, P. C. *Adv. Immun.* **27**, 52-142 (1979).
8. Pfau, C. J., Valenti, J. K., Pevear, D. C. & Hunt, K. D. *J. exp. Med.* **156**, 79-89 (1982).
9. Ahmed, R., Salmi, A., Butler, L. D., Chiller, J. M. & Oldstone, M. B. A. *J. exp. Med.* **160**, 521-540 (1984).
10. Ahmed, R. & Oldstone, M. B. A. *J. exp. Med.* **167**, 1719-1724 (1988).
11. Ahmed, R. et al. *J. Virol.* **62**, 3301-3308 (1988).
12. Salvato, M., Borrow, P., Shimomaye, E. & Oldstone, M. B. A. *J. Virol.* **65**, 1863-1869 (1991).
13. Pircher, H. P., Bürki, K., Lang, R., Hengartner, H. & Zinkernagel, R. *Nature* **342**, 559-561 (1989).
14. Marker, O. & Volkert, M. *J. exp. Med.* **137**, 1511-1525 (1973).
15. Cihak, J. & Lehmann-Grube, F. *Immunology* **34**, 265 (1978).
16. Moskophidis, D., Assmann Wischer, U., Simon, M. M. & Lehmann Grube, F. *Eur. J. Immun.* **17**, 937-942 (1987).
17. Jamieson, B. D., Somasundaram, T. & Ahmed, R. *J. Immun.* **147**, 3521-3529 (1991).
18. Schönrich, G. et al. *Cell* **65**, 293-304 (1991).
19. Rocha, B. & von Boehmer, H. *Science* **251**, 1225-1228 (1991).
20. Schwartz, R. H. *Science* **248**, 1349-1356 (1990).
21. Herman, A., Kappler, J. W., Marrack, P. & Pullen, A. M. *Rev. Immun.* **9**, 745-772 (1991).
22. Webb, S., Morris, C. & Sprent, J. *Cell* **63**, 1249-1256 (1990).
23. Rammensee, H. G., Kroschewski, R. & Frangoulis, B. *Nature* **339**, 541-544 (1989).
24. Ignatowicz, L., Kappler, J. & Marrack, P. *J. exp. Med.* **175**, 917-923 (1992).
25. Saron, M. F., Shidani, B., Nahori, M. A., Guillon, J. C. & Truffa Bachi, P. *J. Virol.* **64**, 4076-4083 (1990).
26. Campbell, I. L., Lepay, D. A. & Oldstone, M. B. A. *J. Cell. Biochem.* **16**, 141 (1992).
27. Jacobs, R. P. & Cole, G. A. *J. Immun.* **117**, 1004-1009 (1976).
28. Odermatt, B., Eppler, M., Leist, T. P., Hengartner, H. & Zinkernagel, R. M. *Proc. natn. Acad. Sci. U.S.A.* **88**, 8252-8256 (1991).
29. Borrow, P., Tishon, A. & Oldstone, M. B. A. *J. exp. Med.* **174**, 203-212 (1991).
30. Tishon, A., Southern, P. J. & Oldstone, M. B. A. *J. Immun.* **140**, 1280-1284 (1988).
31. Zinkernagel, R. M., Leist, T. P., Hengartner, H. & Althage, A. *J. exp. Med.* **162**, 2125-2141 (1985).
32. Doherty, P. C., Zinkernagel, R. M. & Ramshaw, I. A. *J. Immun.* **112**, 1548-1552 (1974).
33. Dunlop, M. B. C. & Blanden, R. V. *J. exp. Med.* **145**, 1131-1143 (1977).
34. Gilden, D. H., Cole, G. A., Monjan, A. A. & Nathanson, N. *J. exp. Med.* **135**, 860-873 (1972).
35. Stitz, L. *Eur. J. Immun.* **22**, 1995-2001 (1992).
36. Peters, M. et al. *Hepatology* **13**, 977-994 (1991).
37. Clerici, M. et al. *J. Immun.* **146**, 2214-2219 (1991).
38. Groux, H. et al. *J. exp. Med.* **175**, 331-340 (1992).
39. Miller, R. G. *Nature* **287**, 544 (1980).
40. Mitchison, N. A. *Proc. R. Soc.* **161**, 275-292 (1964).
41. Pircher, H. P. et al. *Nature* **346**, 629-633 (1990).
42. Pircher, H. P. et al. *Eur. J. Immun.* **22**, 399-404 (1992).
43. Tomonari, K. *Immunogenetics* **28**, 455-458 (1988).

ACKNOWLEDGEMENTS. We thank E. Laine and A. Althage for technical help and H. Hengartner, D. Kagi, U. Kalinke, E. Bucher and M. Battegay for discussions. This work was supported by Swiss National Foundation Grants.

Thymic dendritic cells and T cells develop simultaneously in the thymus from a common precursor population

Carlos Ardavin*, Li Wu, Chung-Leung Li & Ken Shortman†

The Walter and Eliza Hall Institute of Medical Research,
PO Royal Melbourne Hospital, Melbourne, Victoria 3050, Australia

DENDRITIC cells, a minor cell population in lymphoid tissues, are specialized for presentation of antigenic peptides to T lymphocytes¹. Thymic dendritic cells are involved in the deletion of self-reactive T lymphocytes^{2,3}. Although all dendritic cells are ultimately of bone-marrow origin⁴⁻⁷, it has not been clear whether thymic dendritic cells are produced in the adult thymus from a precursor cell or whether they migrate there preformed from the periphery. Recently we isolated from adult mouse thymus a population of early T precursors that could still form B lymphocytes, but not erythroid or myeloid cells, when transferred intravenously^{8,9}. Here we show that these thymic lymphoid precursor cells, as well as bone-marrow haematopoietic stem cells, are able to form both dendritic cells and T-cell progeny when transferred into an irradiated thymus. Such linked development may ensure that developing T cells are negatively selected predominantly by self antigens presented on newly formed thymic dendritic cells.

The minute population of early-T precursors we previously isolated from adult mouse thymus resembled bone marrow haematopoietic stem cells (BMSC) in surface phenotype, except

Received 30 November 1992; accepted 1 February 1993.

1. Mims, C. A. *The Pathogenesis of Infectious Disease* (Academic, London, 1987).
2. Oldstone, M. B. A. *Cell* **56**, 517-520 (1989).
3. Ahmed, R. & Stevens, J. G. in *Virology* (ed. Fields, B. N.) 241-265 (Raven, New York, 1990).
4. Hotchin, J. *Monogr. Virol.* **3**, 1-211 (1971).
5. Lehmann-Grube, F. *Virol. Monogr.* **10**, 1-173 (1971).

* Present address: Department of Cell Biology, Faculty of Biology, Complutense University, 28040 Madrid, Spain.

† To whom correspondence should be addressed.

that they expressed Sca-2 and low levels of CD4 (ref. 8). This 'low-CD4 precursor' population had T-cell antigen receptor (TCR) genes in a germ-line state, produced $\alpha\beta$ and $\gamma\delta$ T cells following intrathymic (i.t.) transfer, but also produced B cells if it seeded bone marrow following intravenous (i.v.) transfer^{8,9}. In a search for the precursors of thymic dendritic cells (DC), we tested these thymic lymphoid precursors, as well as early BMSC¹⁰, because despite the accepted bone-marrow origin of DC⁴⁻⁷, the derivation of DC from the same stem cell as other blood cell lineages had not been proven. These two precursor populations were isolated from adult mice of the *Ly* 5.2 allotype and injected into irradiated congenic *Ly* 5.1 mice. At various times after transfer, DC were released¹¹ from the recipient thymus tissue, then enriched¹¹ so that the very small numbers of DC progeny could be detected. Donor-derived DC were identified as *Ly* 5.2-bearing cells expressing very high levels of class II major histocompatibility complex (MHC) and the DC-specific marker NLDC 145 (refs 11, 12) (Fig. 1).

When 100 purified BMSC¹⁰ were injected into one lobe of an irradiated recipient thymus, 11,000 DC progeny were found in the injected lobe 3 weeks later (Fig. 1), together with 23×10^6 T-lineage cells (as previously shown¹³; Fig. 2). The recipient spleens were also analysed to check for any leakage of BMSC from the injected thymus; no DC progeny (or B-cell or myeloid progeny) were ever found in the spleen after i.t. transfer. DC progeny were found in recipient spleens (Fig. 3), as well as in recipient thymuses, however, after i.v. transfer. These results demonstrate the BMSC origin of both thymic and splenic DC and that the adult thymus can support full DC development.

Intrathymic transfer of the thymic low-CD4 precursor population led to production in the thymus of DC (Fig. 1), as well as T-lineage cells^{8,9} (Fig. 2). The recoverable yield from 10^4 injected low-CD4 precursors was 5×10^6 T-lineage cells and 4×10^3 DC, a 2,000:1 ratio similar to that obtained with BMSC, and similar to the ratio of thymocytes to DC in the normal thymus¹¹. DC

progeny were maintained at this level from 7 to 22 days post-transfer, but they had disappeared from the thymus by 28 days, together with the thymocyte progeny. There is also a rapid turnover of DC in rat thymus¹⁴. DC progeny were produced in the spleen in parallel with thymic DC production when the low-CD4 precursor cells were transferred i.v. (Fig. 3).

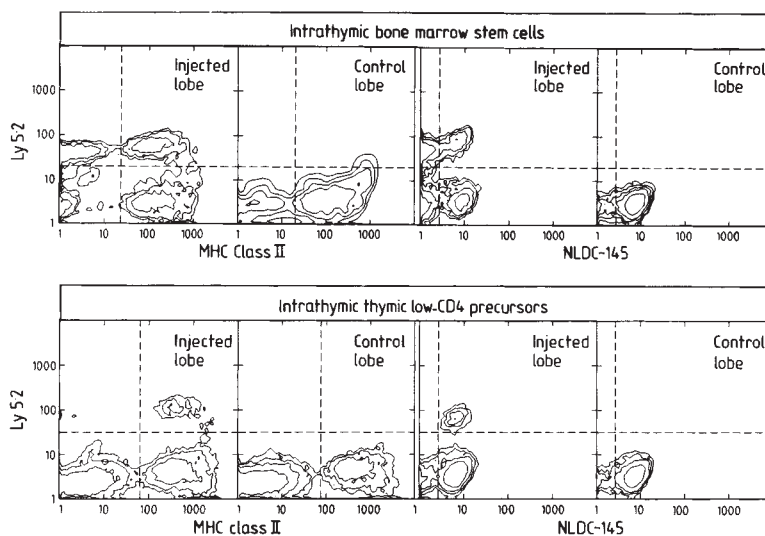
The DC nature of these class II MHC⁺ NLDC-145⁺, *Ly* 5.2⁺ cells in the recipient thymus was confirmed by other criteria (data not shown). These cells showed the high forward and side light-scatter characteristics of DC. They were negative for B220 and low-to-negative for Mac-1, excluding B cells and macrophages that also express class II MHC. They did not express significant levels of TCR- $\alpha\beta$, TCR- $\gamma\delta$, CD3, CD4 or Thy 1. About 60% of them expressed high levels of CD8- α , but only low levels of CD8- β , a characteristic of thymic DC¹¹. Finally, when sorted and incubated at 37 °C in culture medium to allow recovery of normal surface configuration, 90% had typical DC morphology¹¹, with irregular shape and cytoplasmic processes.

Additional separation parameters were used to ensure the DC progeny were derived from the low-CD4 precursor population itself, rather than from contaminating multipotent stem cells or pre-existent DC (data not shown). As the low-CD4 T precursors are Sca-2⁺ and BMSC are Sca-2⁻ (refs 8, 9, 15), the preparation was labelled and sorted for Sca-2 expression; almost all DC precursor activity was in the major Sca-2⁺ fraction. As cells of the low-CD4 precursor preparation are almost entirely class II MHC⁻ (refs 8, 9), they were labelled and sorted to remove any last traces of class II MHC⁺ cells; all DC precursor activity was recovered in the class II MHC⁻ fraction. In addition, cells of the low-CD4 precursor population did not express the DC markers NLDC-145 or N418, and did not have DC morphology.

To determine whether later T precursors with rearranged TCR- γ and - β genes could also form DC, we isolated the CD3⁻4⁻8⁻ IL-2R α ⁺ precursor population from the adult mouse

FIG. 1 The reconstitution of thymic dendritic cells by bone marrow haematopoietic stem cells or by thymic low-CD4 precursor cells. The precursor populations were purified from *Ly* 5.2 mice and injected into the left thymic lobe of irradiated *Ly* 5.1 mice. Fifteen days later thymic DC were prepared from the injected and uninjected lobes and analysed for donor-derived *Ly* 5.2⁺ cells expressing class II MHC and the DC marker NLDC-145. The direct comparison of the injected lobe with the stained cells from uninjected right lobe, used as a negative control, served to correct for background fluorescence when analysing such a minor stromal component in irradiated mice. Note that both the control and the injected lobes of the irradiated recipients contained some host-derived DC, presumably from precursors which survived the irradiation dose. The class II MHC stain was extremely bright, requiring careful colour compensation, and a drop in instrument sensitivity compared to normal staining. This same setting was used for the weaker NLDC-145 stain, resulting in lower fluorescence but also very low background fluorescence. The contour lines in the diagrams are log₂ probability density plots, with 0.2% as the lowest contour. These results at day 15 are typical of 18 separate experiments, with analyses done 7–22 days after transfer.

METHODS. Precursor cells were from 4–6 weeks C57BL/Ka Thy 1.1 (*Ly* 5.2) mice. Early BMSC were isolated¹⁰ by immunomagnetic bead depletion of bone marrow cells bearing mature lineage markers, then sorting for *Ly* 6A/E⁺ cells showing low-to-medium fluorescence after incubation with rhodamine 123. The low-CD4 precursor population was initially isolated from thymus^{8,9} by first depleting thymocytes bearing CD3, CD8, CD2, IL-2R α , Gr-1, TER-119, Mac-1 and B220 (cytotoxic then immunomagnetic bead procedures), then sorting for Thy 1⁺, HSA⁺, H-2K⁺ cells. In later experiments, sorting for class II MHC⁻ was substituted for the class I MHC⁺ parameter, to ensure that there were no preformed DC. The precursor cells (10^2 BMSC, or 10^4 low-CD4 precursors) were injected into the left thymic lobe of γ -irradiated (7.5 Gy) recipient mice (C57BL/6 *Ly* 5.1-Pep^{3b}, 7–8 week)^{8,9}, using 8 recipients per experiment. The thymuses were then removed at various times, injected lobes being pooled for the test and uninjected lobes pooled



as the control. DC were isolated¹¹ by digesting chopped thymic lobes with collagenase, disrupting DC-T cell complexes with ethylene diamine tetraacetate, selecting light-density cells by centrifugation in metrizamide medium, and then depleting T-lineage cells, B cells and macrophages (using antibodies against CD3, CD4, Thy 1, IL-2R α , F4.80, Fc γ , Gr-1 and TER-119, together with anti-mouse-Ig and anti-rat-Ig coated magnetic beads). About 0.1% of thymus cells was recovered. This DC suspension was then stained^{8,9} with fluorescein-conjugated anti-*Ly* 5.2 (clone AL1-4A2), together with biotinylated anti-class II MHC (clone M5/114, specificity I-A^{b,d,q} I-E^{d,k}) or NLDC-145 (ref. 12) (BMA, Augst, Switzerland), using phycoerythrin-avidin as the second stage. Flow cytometric analysis was carried out on a FACScan (Becton Dickinson) using low-angle light scatter and propidium iodide staining to gate out dead cells and debris.

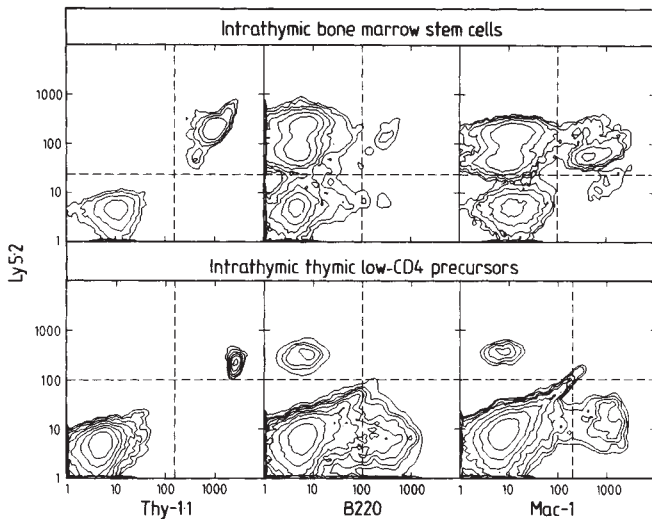


FIG. 2 Reconstitution of thymic T-cell, macrophage and B-cell populations by bone marrow haematopoietic stem cells or by thymic low-CD4 precursor cells. Recipient thymus lobes were analysed 17 days after i.t. transfer of precursor cells. Details are as given in Fig. 1 legend, except that thymus lobes were mechanically disrupted^{8,9} rather than digested, and to analyse T-lineage reconstitution (Thy 1.1 stain), the suspension was counted and stained directly. Note that because the recipients were of the Thy 1.2 allotype, host-derived thymocytes are not visible. Before analysis of macrophage (Mac-1⁺) or B-cell (B220⁺) populations, the suspension was depleted of most thymocytes by treatment with cytotoxic antibodies against CD3, CD4 and CD8, and complement, followed by removal of damaged cells by a metrizamide-density cut^{8,9}; about 0.5% of the thymus cells were recovered. Cells were stained with anti-Ly 5.2, together with lineage-specific biotinylated monoclonal antibodies (anti-Thy 1.1, clone 19-F12; anti-Mac-1, clone M1/70.15; anti-B220, clone RA3-6B2), as for Fig. 1^{8,9}. Analysis was on a FACStar Plus. Contour lines are as in Fig. 1. Results are typical of 4 experiments.

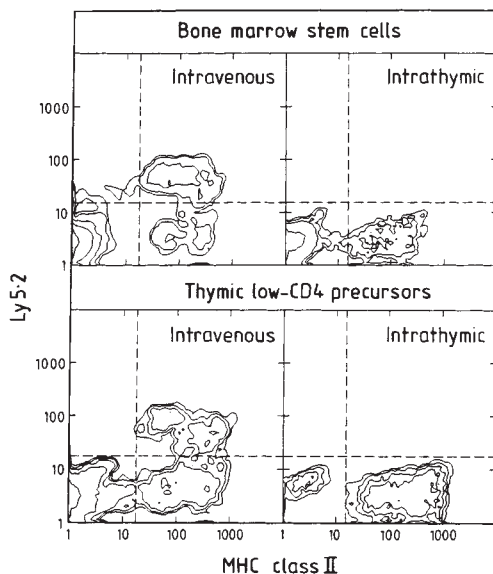


FIG. 3 Reconstitution of splenic dendritic cells by bone marrow haematopoietic stem cells or by thymic low-CD4 precursor cells. Precursor populations were transferred to irradiated recipients by i.t. or i.v. routes, then DC isolated from the recipient spleens 15 days later (as for thymic DC; see Fig. 1) and analysed. Conditions were as outlined for Fig. 1, except that for intravenous transfer 2×10^2 BMSC, or 3×10^4 low-CD4 precursor cells, were injected into recipients that had received two doses of γ -irradiation (each 5.5 Gy, 3 h apart), along with 4×10^4 recipient type bone marrow cells to ensure survival of recipients⁹.

thymus¹⁶. When these were transferred i.t. at a 10-fold higher concentration than the low-CD4 precursors, the levels of T-lineage progeny were as expected¹⁶ but no DC progeny were detected (data not shown).

We considered that thymic low-CD4 precursors could be the origin of all bone-marrow-derived thymus cells, including thymic B cells and macrophages. When 100 BMSC were injected i.t., Mac-1⁺ macrophage progeny were readily detected 15–22 days later, together with a very small amount of B220⁺ B-cell progeny (Fig. 2); the latter were also Ly 6⁺, ThB⁺ and s-Ig⁺, so were true B cells but few in number, agreeing with earlier studies¹³. In contrast, when 10^4 thymic low-CD4 precursors were transferred i.t., no macrophage or B-cell progeny were detected 15–28 days later, although there was extensive T-lineage reconstitution (Fig. 2). This is further evidence for the absence of multipotent haematopoietic stem cells in this preparation and argues for a special developmental relationship between thymic DC and T cells. The inability to generate thymic macrophages is in accordance with the lack of myelopoiesis when thymic low-CD4 precursors are injected i.v.⁹. More surprising is the failure to generate thymic B cells, because normal B-cell progeny are generated in bone marrow and spleen on i.v. transfer⁷. But the thymus is evidently inefficient at supporting B-cell development, and thymic B cells may be a distinct lineage of mainly Ly 1 B cells¹⁷.

Overall, the thymic DC lineage in the adult mouse appears to separate from the T lineage later than the multipotent stem cell but before TCR gene rearrangement. Thymic DC appear to differentiate from an intrathymic precursor, rather than arriving pre-formed as a class II MHC⁺ or NLDC-145⁺ DC, which simply expand in numbers. These thymic DC precursors were indistinguishable from the low-CD4 precursor population representing earliest adult intrathymic precursors of both $\alpha\beta$ and $\gamma\delta$ T cells^{8,9}. But we still lack suitable clonal assays to determine whether this population contains separate but phenotypically similar T-cell and DC precursors, or whether it represents a single type of precursor having both lymphoid and DC developmental potential.

These results have implications for T-cell repertoire selection within the thymus. They suggest that each new cohort of T-lineage cells generated in the thymus is accompanied by (at a 10^3 -fold lower frequency) a parallel cohort of newly generated thymic DC, endogenously derived from a similar if not identical precursor, and having a similar lifespan. This implies that the developing T cells interact mainly with newly formed thymus-restricted DC, rather than with any DC entering from the circulation which might present exogenous foreign antigens (see also ref. 14). This would ensure that the developing T cells are purged only of cells reactive with intrathymic and predominantly self antigens. Other mechanisms¹⁸ would then be needed to produce peripheral tolerance to self antigens not expressed within the thymus itself. □

Received 19 November 1992; accepted 1 March 1993.

- Steinman, R. M. *A. Rev. Immun.* **9**, 271–296 (1991).
- Sprent, J., Lo, D., Gao, E.-K. & Ron, Y. *Immun. Rev.* **101**, 173–190 (1988).
- Fairchild, P. J. & Austyn, J. M. *Int. Rev. Immun.* **6**, 187–196 (1990).
- Steinman, R. M., Lustig, D. S. & Cohn, Z. A. *J. exp. Med.* **159**, 1431–1445 (1974).
- Katz, S. I., Tamaki, K. & Sachs, D. H. *Nature* **282**, 324–326 (1979).
- Pugh, C. M., MacPherson, G. G. & Steer, H. W. *J. exp. Med.* **157**, 1758–1779 (1983).
- Barclay, A. N. & Mayrhofer, G. *J. exp. Med.* **153**, 1666–1671 (1981).
- Wu, L. *et al. Nature* **349**, 71–74 (1991).
- Wu, L., Antica, M., Johnson, G. R., Scollay, R. & Shortman, K. *J. exp. Med.* **174**, 1617–1627 (1991).
- Li, C.-L. & Johnson, G. R. *J. exp. Med.* **175**, 1443–1447 (1992).
- Vremec, D. *et al. J. exp. Med.* **176**, 47–58 (1992).
- Kraal, G., Breel, M., Janse, M. & Bruin, G. *J. exp. Med.* **163**, 981–997 (1986).
- Spangrude, G. J. & Scollay, R. *J. Immun.* **145**, 3661–3668 (1990).
- Kampinga, J., Nieuwenhuis, P., Roser, B. & Aspinall, R. *J. Immun.* **145**, 1659–1663 (1990).
- Spangrude, G. J., Heimfeld, S. & Weissman, I. L. *Science* **241**, 58–62 (1988).
- Pearse, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 1614–1618 (1989).
- Than, S., Inaba, K., Fukuda, Y., Adachi, Y. & Ikehara, S. *Eur. J. Immun.* **22**, 1299–1301 (1992).
- Miller, J. F. A. P. & Morahan, G. *A. Rev. Immun.* **10**, 51–69 (1992).

ACKNOWLEDGEMENTS. We thank R. Scollay, D. Metcalf and G. Johnson for discussion. This work was supported by the National Health and Medical Research Council, Australia, and the US Army Medical Research and Development Command, C.A. received an ICRET Fellowship from the International Union against Cancer.

A new subunit of the cyclic nucleotide-gated cation channel in retinal rods

T.-Y. Chen*, Y.-W. Peng*, R. S. Dhallan†, B. Ahamed‡, R. R. Reed*§ & K.-W. Yau*‡||

‡ Howard Hughes Medical Institute, and Departments of * Neuroscience, † Biomedical Engineering, and § Molecular Biology and Genetics, The Johns Hopkins School of Medicine, 725 North Wolfe Street, Baltimore, Maryland 21205, USA

RETINAL rods respond to light with a membrane hyperpolarization produced by a G-protein-mediated signalling cascade that leads to cyclic GMP hydrolysis and the consequent closure of a cGMP-gated channel that is open in darkness¹⁻⁷. A protein that forms this channel has recently been purified from bovine retina⁸ and molecularly cloned⁹, suggesting that the native cGMP-gated channel might be a homo-oligomer¹⁰. Here we report the cloning of another protein from human retina which has only about 30% overall identity to the rod channel subunit. This protein, immunocytochemically localized to rod outer segments, does not form functional channels by itself. However, when co-expressed with the cloned human rod channel protein¹¹, it introduces rapid flickers to the channel openings that are characteristic of the native channel¹²⁻¹⁶. The hetero-oligomeric channel is also highly sensitive to the blocker L-cis-diltiazem, like the native channel^{15,17,18}. This new protein thus seems to be another subunit of the native rod channel. The hetero-oligomeric nature of the rod channel means that it is no exception to a common motif shared by other ligand-gated channels.

The coding region of the rod cGMP-gated channel from human retina¹¹, here called hRCNC1, was used to screen a human retinal cDNA library (Fig. 1 legend). Two new classes of clones, referred to as hRCNC2a and hRCNC2b, were identified that differed only at their 5' ends. Partial genomic cloning has confirmed that they represent alternatively spliced products of a single gene, with the 5' end of hRCNC2b being derived from one or more exons situated upstream of those for hRCNC2a. Nucleotide sequence analysis indicates, however, that the alternatively spliced 5' end of hRCNC2a is untranslated; thus, hRCNC2a gives rise to a protein that is identical to hRCNC2b except for a missing N-terminal region. The proteins encoded by the two messages are, respectively, 623 and 909 amino acids long (*M_r* 70,843 and 102,330, respectively), versus 686 amino acids for hRCNC1 (Fig. 1a). The two proteins show only about 30% overall identity to hRCNC1 in regions of overlap, but the cyclic nucleotide-binding domain⁹ (50% identity) and the so-called S4 region^{10,19} are clearly recognizable. Furthermore, hydrophobicity plots of these proteins (only hRCNC2a shown in Fig. 1b) superimposed on that of hRCNC1 indicate that molecular foldings in the cell membrane are comparable, although the putative H4 transmembrane domain^{9,10} in hRCNC1 (see also ref. 20) is not obvious in these proteins. As the transmembrane topology of hRCNC1 is still uncertain, the significance of this hydrophobicity difference remains to be clarified.

A rabbit polyclonal antibody (JH486) was raised against the C-terminal peptide present in hRCNC2a and hRCNC2b (Fig. 2 legend). In a human retinal section, this antibody stains all of the rod (but not cone) outer segments, plus weaker staining of the outer plexiform layer (Fig. 2a). A second antibody (JH487) raised against an N-terminal peptide present in hRCNC2b but not hRCNC2a (Fig. 2 legend) stained the outer plexiform layer but not the rod outer segments (Fig. 2b). Pread-

sorption of the respective peptides to these antibodies abolished all staining, indicative of the specificity of the immunoreactivities. These results suggest that hRCNC2a is located predominantly, perhaps exclusively, in the rod outer segment, whereas hRCNC2b may be present in the outer plexiform layer.

These results together point to the existence of a new subunit of the native rod channel. Henceforth we focus on hRCNC2a because this appears to be the alternatively spliced variant in rod cells. Using transient transfection in the human embryonic kidney 293 cell line followed by excised-patch recording (Fig. 3 legend), we have been unable to demonstrate any ability of hRCNC2a to form functional channels by itself. By contrast, when co-transfected with hRCNC1, hRCNC2a produced cGMP-gated channel activity with kinetics unlike those of hRCNC1 expressed alone. Figure 3a (left panel) shows cGMP-induced channel openings from a cell transfected with only hRCNC1. In agreement with previous work¹¹, these channel openings had a conductance of 25–30 pS (sub-conductances were rarely seen), and they showed perhaps three distributions of open time, of the order of 1 ms, several milliseconds, and tens of milliseconds, respectively. The long openings, with their stable rectangular shapes, were not observed in native channels of amphibian^{12-14,16} or mammalian¹⁵ rod outer segments (but see ref. 16), which instead show bursts of flickery activity consisting of openings lasting ~1 ms or less. Expression of the cloned bovine rod cGMP-gated channel subunit in *Xenopus* oocytes⁹ also produced long, stable openings. Figure 3a (right panel), by contrast, shows cGMP-induced channel openings from a cell co-transfected with hRCNC1 and hRCNC2a. The activity in this case is clearly different, with the long openings chopped into flickers like those seen in recordings from rod outer segments¹²⁻¹⁶; there was also an indication of the existence of sub-conductance states associated with the flickering. It thus seems that the prominent flickery kinetics of the native channel arise from this new channel subunit.

We have examined the macroscopic current-voltage relation for the channel at a saturating cGMP concentration. Without divalent cations, the influence from hRCNC2a is small (Fig. 3b). With divalent cations on the extracellular side of the membrane, the outward rectification resulting from divalent cation blockage becomes less severe in the presence of hRCNC2a (Fig. 3c); a quantitative comparison with results from amphibian

FIG. 1 a, Amino-acid sequences for hRCNC2a and hRCNC2b deduced from their cloned cDNA nucleotide sequences and aligned with the sequence for hRCNC1. The methionine marked by an asterisk indicates the first amino acid in the hRCNC2a sequence; amino acids N-terminal to this methionine (286 altogether) are present only in hRCNC2b. Numbering of amino acids is for hRCNC2a. In the alignment (using the Wisconsin Package), a hyphen means the absence of an amino acid, and vertical dashes indicate identical amino acids. H1 to H5 are putative transmembrane domains for hRCNC1 (refs 9, 10 and 20). S4 corresponds to the voltage-sensing domain in voltage-gated ion channels, characterized by the presence of arginine or lysine in every third or fourth position^{10,19}. P indicates the putative pore region for hRCNC1, identified by its homology with that in the Shaker potassium channel³⁰⁻³³. cG-binding site indicates binding domain for cGMP^{9,10}. b, Comparison of hydrophobicity plots of hRCNC2a and hRCNC1. Solid profile corresponds to hRCNC2a, dashed profile to hRCNC1. Gaps on the amino-acid scales are introduced to match up with the alignments in a. The H1-5, S4, P and cG symbols have the same meanings as in a. The hydrophobicity index is calculated by the method of Kyte and Doolittle³⁴, with positive values indicating hydrophobic regions. METHODS. A partial cDNA clone of hRCNC1 was used as a probe to screen ~5 × 10⁵ recombinants from an adult human retinal cDNA library³⁵ at both low (55 °C, 2 × SSC wash) and high (65 °C, 0.5 × SSC wash) stringency³⁶. Those clones that hybridized at low but not at high stringency were assumed to be new ones and were analysed further. One of length ~1 kb had a region of strong homology with the putative cyclic nucleotide-binding domain in hRCNC1. To obtain a full-length clone, an oligonucleotide derived from the 5' end of this clone was used as a probe to rescreen 2 × 10⁶ recombinants from the same library. Two of the longest hybridizing clones correspond to hRCNC2a and hRCNC2b.

|| To whom correspondence should be addressed.

a

hRCNC2 MPRELSRIEEKEDEEEEEEEEEEEVTEVLLDSCVVSQVGVGQSEEDGTRPQSTSDQKLWEEVGEEAKKEAEKAK
hRCNC1 -----

hRCNC2 EEAEVAAEEAAKEPQDWAETKEEPEAAEAASSGVPA TKQH---PEVQVEDTDADSCPLMAEENPPSTVLPPSPAKS
hRCNC1 1 -----MKNNIIINTQQS FVTMPNVIVPDIEKEIRRM---ENGAC----- 35

hRCNC2 DTLIVPSSASGTHRKKLPSEDEAEELKALSPAESPVVAWSDPTTPKDTDGQDRAASTASTNSAIINDRLQELVKLFKER
hRCNC1 36 -----SSFSEDDDSASTSESEENENPHARGSFYSKSL-RKGGPSQREQYLPGAIALFNVNSSNKDQEPEEKKKKKKK 108

hRCNC2 1 TEKVKEKLIDPDVTSDEESPKPSPAKKAPEPADTKPAEAEPEVEEHYCDMLCKFKHRPWKKYQFPQSIDPLTNLMYVL 30
hRCNC1 109 KSKSDKNKNENKNDPEKKKKKKDKKKKKKKSKDKKEEKKEV-----VVIDPSGNTYYN- 163

hRCNC2 31 WLFFVVMAWNWCWLPVRWAFPYQTPDNIHWWLMDYLCDLIYFLDITVFQTRLQFVRGGDIITDKDMRNNYLSRRF 110
hRCNC1 164 WLFCTITLPMYNTWTVIARACFDELQSDYLEYWLILDYVSDIVYLIDMFV-RTRTGYLEQGLLVKEELINKYKSNLQF 242

hRCNC2 111 KMDLLSLPLDLYLKVGVN-PLRLPRCLKYMAFFEFNSRLESILSKAYVYRVIRTAYLLYSLHLNSCLYIYASAYQG 189
hRCNC1 243 KLDVLSLIPTDLYLKVGVN-PLRLPRCLKYMAFFEFNSRLESILSKAYVYRVIRTAYLLYSLHLNSCLYIYASAYQG 322

hRCNC2 190 LGSTHWVYDGVGNS-----YIRCYYFAVKTLITIGGLPDKTLFEIVFOLLNYFTGVFAFVSMIGQMRDVGGAATAG 261
hRCNC1 323 FGNDTWVYPDINDPEFGLRARKYVYSLYWSTLTLTITIGETPPPPVRDSEYVFFVVDPLIGVLIFATIVGNIGSMISNMNAA 402

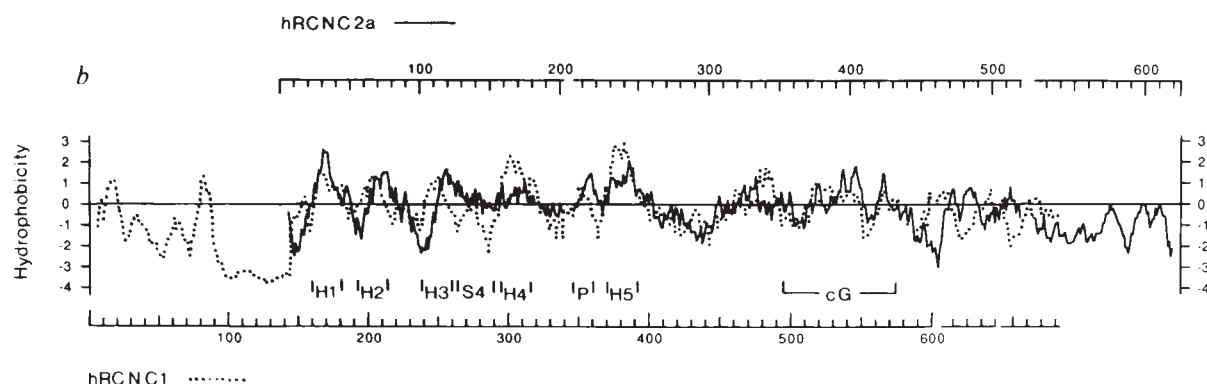
hRCNC2 262 QTYIRSCHMDSTVKYMFYKIPKSVQNRVKTWYEYTWHSQGLDESELMVQLPDKMRDLDAIDVNYNIVSKVALFQGC DRQ 341
hRCNC1 403 RAEFQARIDAIKQYMHFRNVSKDMEKRVIKWFDYLTWNTKKTVDKEVLKYLDPDKLRAEIAINVHLDLTKKVRIFADCEAG 482

hRCNC2 342 MIFDMLKRLRSVVYLPNDYVCKKGEIGREMYIIQAGQVQLGGPDGKSVLVTLKAGSVFGEISLLAVGG---GNRRTANV 418
hRCNC1 483 LLVELVLKLPQVYSPGDYICKKGDIGREMYIIKEGKLA VVAD-DGVTQFVVLSDGSYFGEISILNIKSGKAGNRRTANI 561

hRCNC2 419 VAHGFTNLFIIDKKDLNEILVHPESQKLLRKKARRMLRSNNKPKEKSVLILPPRAGTPKLFNAALAMTGKMGKGAKG 498
hRCNC1 562 KSIGYSDLFCLSKDDLMEALTEYPAKTMLEEKGKQILM-----KDGLLDLNIANAGSDPKDLEEKV---TRMEGSVDLLQ 634

hRCNC2 499 GKLAHLRLRLKELAALEAAAKQ-----QELVEQAKSSQDVKEEGSAAPDQHTHPKEAATDPPAPRTPPEPPGSP 569
hRCNC1 635 TRFARI---LAYESMQQKQLKRLTKVEKFLKPLIDTEFSSIEGPGAESGPIDST----- 686

hRCNC2 570 SSPPPASLGRPEGEEGPAEPEEHSVRICMSPGPEGEQILSVKMPEEREKAE 623
hRCNC1 -----



rods^{21,22}, the only preparation studied under the same conditions, also suggests that the hetero-oligomeric channel may be closer to the native channel in this respect. Finally, Fig. 3*d* compares the cGMP dose-response relations for the expressed homo-oligomeric and hetero-oligomeric channels. The influence of hRCNC2a is again slight: in its absence or presence, the dose-response relation shows a Hill coefficient of near 2 and $K_{1/2}$ of 60–80 μM at +60 mV and near 100 μM at –60 mV, broadly similar to the native amphibian¹ or mammalian^{15,23} rod channel.

We have also examined the effect of the blocker *L-cis*-diltiazem, which inhibits the cyclic nucleotide-gated current in mammalian^{15,24} and in amphibian^{17,18} rods at micromolar concentrations. For the expressed channel consisting of hRCNC1 alone, the K_i for diltiazem was near 100 μM at +60 mV and higher at negative voltages (Fig. 4*a*). For the expressed channel consisting of both hRCNC1 and hRCNC2a, however, the K_i for diltiazem decreased to ~1 μM at +60 mV and 10 μM at –60 mV (Fig. 4*b*), values very similar to those for the native channel in bovine rods under the same recording conditions¹⁵. There is also a characteristic time-dependence in the diltiazem block of the native bovine channel¹⁵ which was reproduced here by the presence of hRCNC2a (Fig. 4*b*, upper panel). Thus, hRCNC2a seems to be important for the binding of *L-cis*-diltiazem. These findings provide further evidence that hRCNC2a is an integral subunit of the native channel.

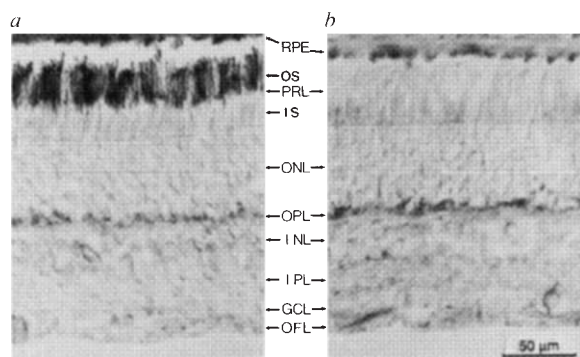


FIG. 2 Immunoperoxidase staining of an 8 μm section of human retina with an anti-peptide antibody against both hRCNC2a and hRCNC2b (*a*) and against hRCNC2b alone (*b*), using Nomarski differential interference contrast optics. The anatomical layers are as follows: RPE, retinal pigment epithelium; PRL, photoreceptor layer (containing the outer segments (OS) and the inner segments (IS) of the receptor cells); ONL, outer nuclear layer (containing the cell bodies of the photoreceptors); OPL, outer plexiform layer; INL, inner nuclear layer; IPL, inner plexiform layer; GCL, ganglion cell layer; OFL, optic fibre layer.

METHODS. The antibodies are rabbit polyclonal antibodies commercially generated (Hazleton, PA) against synthetic peptides coupled to bovine serum albumin. JH486 is against the peptide YQLSVKMPPEEREKAE, found at the C-terminal of both hRCNC2a and hRCNC2b; the tyrosine residue was added for the purpose of peptide purification. JH487 is against the peptide AESPVVAWSDPTTPK, present only near the N-terminal of hRCNC2b. Immunostaining was on frozen retinal sections prefixed with 4% paraformaldehyde³⁷. Sections were first incubated with 5% normal goat serum (Vector Lab.) in phosphate-buffered saline (PBS) for 1 h at room temperature to reduce background staining. They were then incubated with the primary antibody (1:5,000 dilution) overnight at 4 °C, followed by two washes in PBS for 30 min. Triton X-100 (0.3%) was added to all incubation and wash buffers. Sections were next incubated with a biotin-conjugated goat anti-rabbit secondary antibody (Vector Lab; 1:200 dilution) for 2 h at room temperature, then washed twice in PBS for 30 min, followed by a 1-h incubation with an avidin-biotin-peroxidase complex (Vector Lab; 1:100 dilution) in PBS. After two more washes for 30 min, the stain was developed with a substrate solution of 20 ml PBS, 0.1 ml 3% H_2O_2 and 10 mg diaminobenzidine. The staining reaction was terminated by washing with PBS, and the sections were coverslipped with 50% glycerol in PBS. In the preadsorption experiments, the primary antibody was incubated with the corresponding synthetic peptide in a peptide:antibody ratio of 1 mg:1 ml for at least 48 h at 4 °C before dilution and staining.

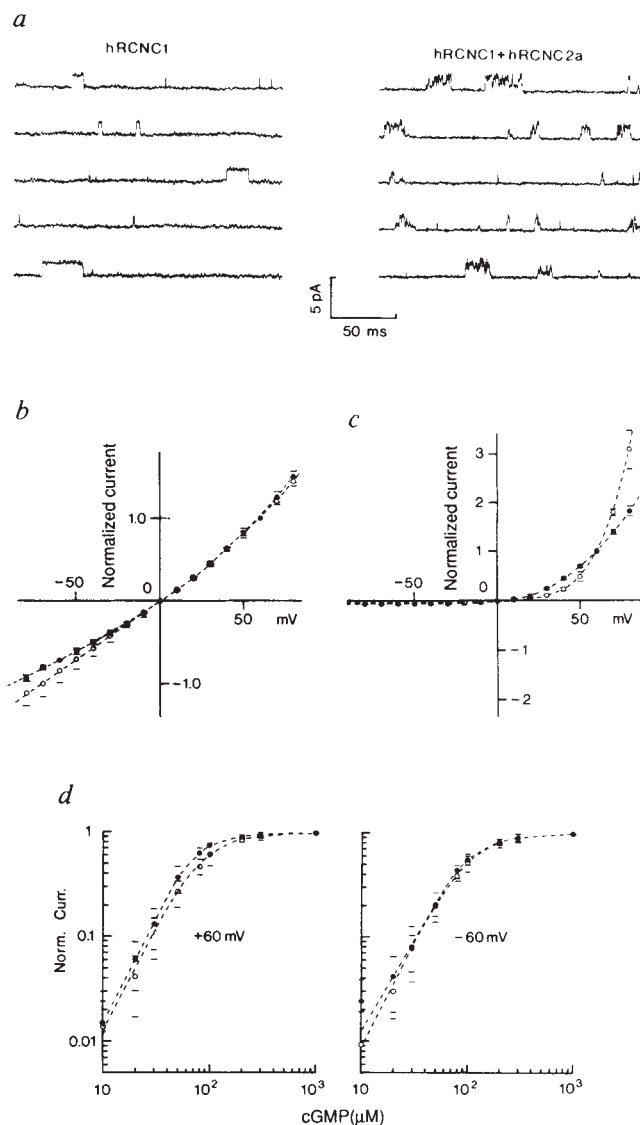


FIG. 3 Comparisons between cGMP-gated channels produced by expression of hRCNC1 alone and expression of hRCNC1 and hRCNC2a together. All results were obtained with electrical recordings from inside-out patches of plasma membrane excised from 293 human embryonic kidney cells transfected with the respective cDNA. *a*, Single-channel openings induced by 20 μM cGMP at +60 mV (left and right panels); DC-2kHz bandwidth. Traces have been selected to show the different open times or burst durations. Symmetrical solutions were without divalent cations (see below). *b*, Macroscopic current-voltage relation induced by saturating (1,000 μM) cGMP in the absence of divalent cations. Leakage current is already removed. cGMP-induced current is normalized with respect to the value at +60 mV. Filled circles, hRCNC1+hRCNC2a (average of 5 experiments); open circles, hRCNC1 alone (average of 8 experiments). *c*, Macroscopic current-voltage relation induced by saturating (1,000 μM) cGMP in the presence of extracellular divalent cations (see below). Leakage current again removed, and current normalized to unity at +60 mV. Filled circles, hRCNC1+hRCNC2a (average of 4 experiments); open circles, hRCNC1 alone (average of 5 experiments). *d*, Dose-response relations at ± 60 mV in the absence of divalent cations. Filled circles, hRCNC1+hRCNC2a (average of 5 experiments at each voltage); open circles, hRCNC1 alone (average of 5 experiments at each voltage). Current is normalized with respect to value at 1,000 μM cGMP. In *b*, *c* and *d*, horizontal bars indicate standard deviations and all dashed curves are fitted to points by eye.

METHODS. For transient transfections³⁸, a pCIS expression vector was used²⁸. Co-transfections were done with a 1:1 ratio of hRCNC1 and hRCNC2a. Patch-clamp recordings were made at 2–4 days after transfection, with borosilicate glass electrodes having tip lumen diameters of ~1 μm . Solutions were: with divalents, 140 mM NaCl, 5 mM KCl, 10 mM Na-HEPES, pH 7.6, 2 mM CaCl_2 , 1 mM MgCl_2 ; without divalents, 140 mM NaCl, 5 mM NaCl, 5 mM KCl, 10 mM Na-HEPES, pH 7.6, 0.5 mM Na-EGTA, 0.5 mM Na-EDTA.

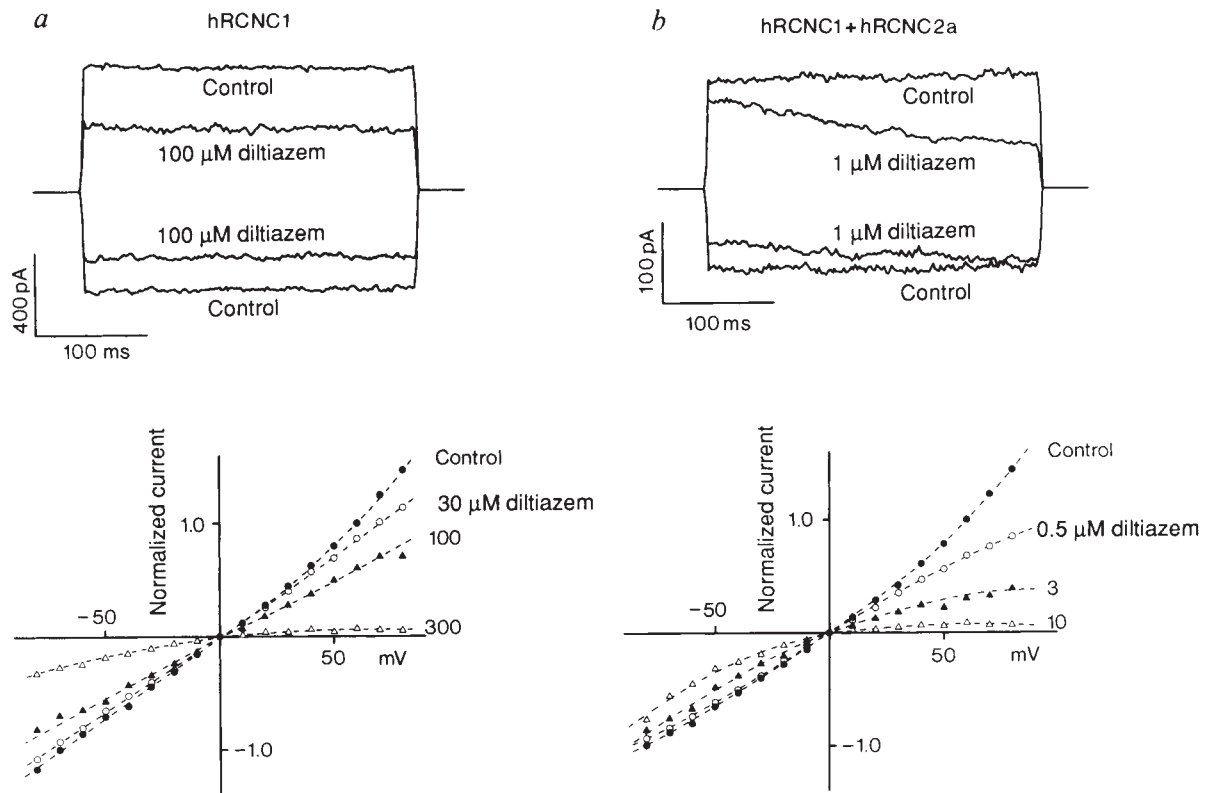


FIG. 4 Effect of *L-cis*-diltiazem on the expressed cGMP-gated channel. *a*, hRCNC1; *b*, hRCNC1+hRCNC2a. In both cases the upper panel shows an experiment in which the membrane potential was stepped from zero to ± 80 mV in the presence of $500 \mu\text{M}$ cGMP minus (control trace) or plus diltiazem at the indicated concentrations. Lower panels show averaged results from many experiments ($n=2-5$ for each diltiazem concentration including control) covering the full voltage range. Current amplitudes were

measured at the end of the 300 ms voltage pulse, when the current in the presence of diltiazem was near steady state. Dashed curves were fitted by eye. Note the more effective blockage at positive voltages, and a time-dependence of the blockage for the hetero-oligomeric channel in *b*. Symmetrical solutions were without divalent cations, with diltiazem present on the cytoplasmic side of the membrane.

In summary, we have identified what appears to be a new subunit of the cGMP-gated channel in retinal rods. This subunit can account for the highly characteristic flickery open-channel activity observed in native rod membrane. The exact physiological purpose of this feature, however, is unclear, although it has also been observed in an ATP-activated channel in dorsal root ganglion neurons²⁵. In addition, based on the observation shown in Fig. 3c, there may be an influence from hRCNC2a on the permeability of the native channel to divalents, which is an important feature in phototransduction¹⁻⁷. At present we do not know the significance of the alternative splicing that produces the hRCNC2a and hRCNC2b variants, especially because hRCNC2b produces similar channel behaviour to hRCNC2a when co-transfected with hRCNC1 (data not shown). It will be interesting if hRCNC2b is indeed present in the retinal outer plexiform layer, considering reports of a putative cGMP-gated channel in on-bipolar cells^{26,27}.

Questions that arise with the emergence of the new subunit described here concern the size of the family of cyclic nucleotide-gated channel subunits and whether the cone cGMP-gated channel and the olfactory cAMP/cGMP-gated channel are also hetero-oligomers. At least for the latter this may well be the case, because the expressed rat olfactory channel²⁸ is unlike the native channel²⁹ in its discrimination between cAMP and cGMP. □

4. Stryer, L. *J. biol. Chem.* **266**, 10711-10714 (1991).
5. Yau, K.-W. *Curr. Opin. Neurobiol.* **1**, 252-257 (1991).
6. Kaupp, U. B. & Koch, K.-W. *A. Rev. Physiol.* **54**, 153-173 (1992).
7. Lagnado, L. & Baylor, D. A. *Neuron* **8**, 995-1002 (1992).
8. Cook, N. J., Hanke, W. & Kaupp, U. B. *Proc. natn. Acad. Sci. U.S.A.* **84**, 585-589 (1987).
9. Kaupp, U. B. *et al. Nature* **342**, 762-766 (1989).
10. Kaupp, U. B. *Trends Neurosci.* **14**, 150-157 (1991).
11. Dhallan, R. S. *et al. J. Neurosci.* **12**, 3248-3256 (1992).
12. Haynes, L. W., Kay, A. R. & Yau, K.-W. *Nature* **321**, 66-70 (1986).
13. Zimmerman, A. L. & Baylor, D. A. *Nature* **321**, 70-72 (1986).
14. Matthews, G. & Watanabe, S.-I. *J. Physiol., Lond.* **403**, 389-405 (1988).
15. Quandt, F. N., Nicol, G. D. & Schnetkamp, P. P. M. *Neuroscience* **42**, 629-638 (1991).
16. Torre, V., Straforini, M., Sesti, F. & Lamb, T. D. *Proc. R. Soc. B250*, 209-215 (1992).
17. Stern, J. H., Kaupp, U. B. & MacLeish, P. R. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1163-1167 (1986).
18. Haynes, L. W. *J. gen. Physiol.* **100**, 783-801 (1992).
19. Jan, L. Y. & Jan, Y. N. *Nature* **345**, 672 (1990).
20. Wohlfart, P., Haase, W., Molday, R. S. & Cook, N. J. *J. biol. Chem.* **267**, 644-648 (1992).
21. Yau, K.-W., Haynes, L. W. & Nakatani, K. in *Membrane Control of Cellular Activity* (ed. Lüttgau, H. C.) 343-366 (Fischer, Stuttgart, 1986).
22. Matthews, G. *J. Neurosci.* **6**, 2521-2526 (1986).
23. Lühning, H., Hanke, W., Simmoteit, R. & Kaupp, U. B. in *Sensory Transduction* (eds Borsellino, A., Cervetto, L. & Torre, V.) 169-173 (Plenum, New York, 1990).
24. Koch, K.-W. & Kaupp, U. B. *J. biol. Chem.* **260**, 6788-6800 (1985).
25. Bean, B. P., Williams, C. A. & Ceelen, P. W. *J. Neurosci.* **10**, 11-19 (1990).
26. Nawy, S. & Jahr, C. E. *Nature* **346**, 269-271 (1990).
27. Shiells, R. A. & Falk, G. *Proc. R. Soc. B247*, 21-25 (1992).
28. Dhallan, R. S., Yau, K. W., Schrader, K. A. & Reed, R. R. *Nature* **347**, 184-187 (1990).
29. Frings, S., Lynch, J. W. & Lindemann, B. *J. gen. Physiol.* **100**, 45-68 (1992).
30. MacKinnon, R. & Yellen, G. *Science* **250**, 276-279 (1990).
31. Yellen, G., Jurman, M. E., Abramson, T. & MacKinnon, R. *Science* **251**, 939-944 (1991).
32. Yool, A. J. & Schwarz, T. L. *Nature* **349**, 700-704 (1991).
33. Hartmann, H. A. *et al. Science* **251**, 942-944 (1991).
34. Kyte, J. & Doolittle, R. F. *J. molec. Biol.* **157**, 105-132 (1982).
35. Nathans, J., Thomas, D. & Hogness, D. S. *Science* **232**, 193-202 (1986).
36. Jones, D. T. & Reed, R. R. *Science* **244**, 790-796 (1989).
37. Peng, Y. W., Sharp, A. H., Snyder, S. H. & Yau, K.-W. *Neuron* **6**, 525-531 (1991).
38. Gorman, C. M., Gies, D. R. & McCray, G. *DNA and Protein Engineering Techniques* **2**, 3-9 (1990).

Received 11 January; accepted 16 February 1993.

1. Yau, K.-W. & Baylor, D. A. *A. Rev. Neurosci.* **12**, 289-327 (1989).
2. Pugh, E. N. & Lamb, T. D. *Vision Res.* **30**, 1923-1948 (1990).
3. McNaughton, P. A. *Physiol. Rev.* **70**, 847-883 (1991).

ACKNOWLEDGEMENTS. We thank L. W. Haynes, T. Kurahashi and G. Yellen for technical advice; Y. Koutalos, R.-C. Huang, and T. Kurahashi for comments on the manuscript; and Y. Koutalos for suggesting the diltiazem experiment. *L-cis*-diltiazem was a gift from Tanabe Seiyaku Co. (Osaka, Japan). This work was supported by a grant from the US National Eye Institute to K.-W.Y.

Mouse microcytic anaemia caused by a defect in the gene encoding the globin enhancer-binding protein NF-E2

Luanne L. Peters*, Nancy C. Andrews*,
Eva M. Eicher†, Mark B. Davidson*,
Stuart H. Orkin*‡ & Samuel E. Lux*§

* Division of Hematology/Oncology, Children's Hospital and Dana Farber Cancer Institute, Department of Pediatrics, Harvard Medical School, Boston, Massachusetts 02115, USA

† The Jackson Laboratory, Bar Harbor, Maine 04609, USA

‡ Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA

THE nuclear DNA-binding protein NF-E2 is thought to mediate the powerful erythroid enhancer activity of the α - and β -globin locus control regions¹⁻⁴ and participates in the control of genes encoding two enzymes of haem biosynthesis (prophobilinogen deaminase and ferrochelatase)^{1,5}. The major component of NF-E2 is a 45K polypeptide (designated p45 NF-E2) that belongs to the basic region-leucine zipper family of transcription factors⁶. This subunit of NF-E2 is specifically expressed in haematopoietic progenitor cells and differentiated cells of the erythroid, megakaryocyte and mast cell lineages⁶. The gene encoding p45 NF-E2 (murine gene *Nfe2*) has been mapped to mouse chromosome 15 near the mutation microcytosis (*mk*). Homozygous *mk* mice have severe hypochromic microcytic anaemia as a result of decreased globin synthesis and defects in intestinal and erythroid iron absorption. Here we investigate whether the *mk* mutation lies within *Nfe2* by characterizing the p45 NF-E2 gene and determining its DNA sequence in wild-type and *mk* alleles. The *mk* allele carries a missense mutation that causes substitution of valine by alanine at amino acid 173 of the p45 NF-E2 protein. Expression of p45 NF-E2 messenger RNA was detected in erythroid tissues of normal mice and in the duodenum of normal and severely anaemic β -thalassaemic (*Hbb*^{d-th3}/*Hbb*^{d-th3}) mice. We propose that the *mk* mutation results in an impaired form of NF-E2 which fails to regulate both globin production and iron metabolism properly.

We used a 1.5-kilobase (kb) complementary DNA fragment to map the p45 NF-E2 gene in the mouse by Southern blot analysis of DNA from the C57BL/6J $\times$ DBA/2J (B $\times$ D) recombinant inbred (RI) strain set and an interspecific backcross. Analysis of B $\times$ D RI lines localized *Nfe2* to within one map unit of the *Spt-2* (salivary protein-2) locus on mouse chromosome 15 (Fig. 1). By Southern blot analysis of 77 interspecific (BALB/cJ $\times$ CAST/Ei) backcross progeny, 6 recombinants between *Nfe2* and the *Gdc-1* (glycerol phosphate dehydrogenase-1) locus were obtained, placing *Nfe2* within eight map units of *Gdc-1*, which is located on the distal third of chromosome 15.

The autosomal recessive mutation designated microcytosis (*mk*) lies approximately one map unit proximal to *Spt-2*. The *mk* mutation causes severe hypochromic microcytic anaemia. Adult *mk/mk* mice have increased numbers of red blood cells with a low mean corpuscular volume, decreased (but balanced) globin chain synthesis, decreased haemoglobin and haematocrit, elevated reticulocyte counts, and splenomegaly^{7,8}. Uptake of iron by *mk/mk* red blood cells is decreased relative to normal, and iron uptake by *mk/mk* intestinal mucosa is defective, resulting in depleted iron stores in *mk/mk* tissues⁸⁻¹¹. The anaemia and the depleted iron stores are not ameliorated in *mk/mk* mice by oral iron therapy, but a partial response to parenteral iron administration occurs^{10,11}. Bone marrow transplantation experi-

ments indicate that the *mk* phenotype results from defects intrinsic to both haematopoietic stem cells and intestinal cells¹². An apparently similar familial disorder of unknown aetiology has been described in humans¹³.

To investigate whether *mk* represents a mutation at the *Nfe2* locus, we determined the genomic structure of the wild-type gene for p45 NF-E2 and then sequenced the exons from the *mk* allele. The *Nfe2* gene contains three exons spanning ~7.5 kb (Fig. 2). Exon 1 contains 5'-untranslated sequences. The translation initiation codon lies in exon 2. The DNA-binding and dimerization domains of the protein are encoded in exon 3.

To interpret potential sequence differences between wild-type and *mk* alleles, it was necessary to investigate the strain origin of the wild-type *Nfe2* allele that gave rise to the *mk* mutation. The *mk* mutation was originally observed in (C57BL/6J $\times$ DBA/2J) F₂ mice. Surviving *mk/mk* mice were crossed to normal C3H/HeJ mice to produce obligate heterozygotes. The inbred MK/RE-*mk*/+ strain was then established by repeated intercrossing of sibs derived from these animals⁸. Hence, if *mk* is a mutation in the *Nfe2* gene, a mutant form of one of the original parental (C57BL/6J or DBA/2J) *Nfe2* alleles remains

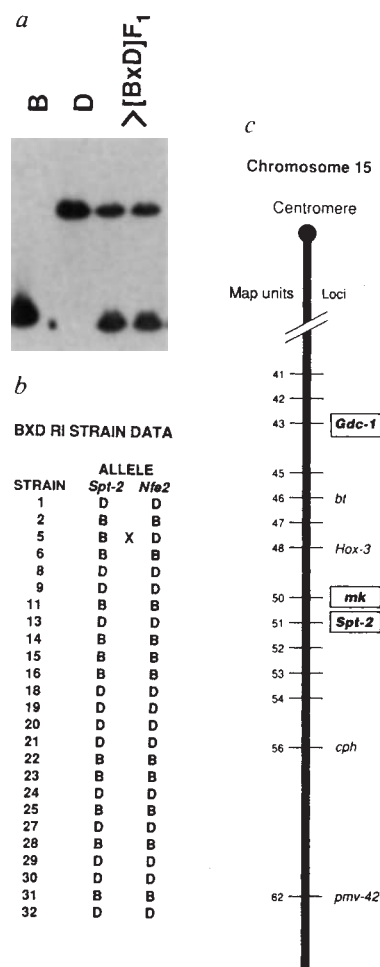


FIG. 1 Mapping of *Nfe2* using the BXD RI strains. **a**, A *Hind*III restriction fragment length polymorphism was detected between parental C57BL/6J (2.0-kb fragment) and DBA/2J (6.0-kb fragment) strains, designated B and D, respectively, using a 1.5-kb p45 NF-E2 cDNA that includes the entire coding sequence. As expected, both fragments are present in F₁ progeny derived from crosses between C57BL/6J and DBA/2J (**b**). The strain distribution pattern obtained for *Nfe2* and *Spt-2* in each of the 25 B $\times$ D RI strains is shown. One recombinant was detected (indicated by 'X'), placing *Nfe2* within 1 map unit of *Spt-2* on chromosome 15 (95% confidence limits 0.03–7.32 maps units). **c**, Map of the distal segment of mouse chromosome 15 showing the positions of *Spt-2*, *Gdc-1* and *mk*.

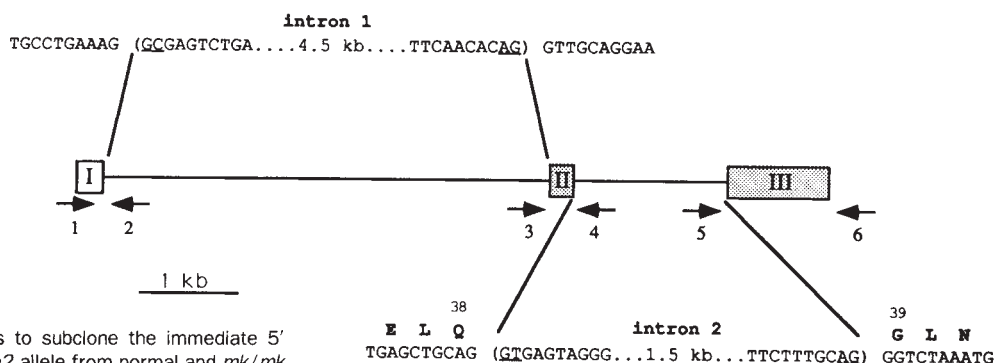
§ To whom correspondence should be addressed.

FIG. 2 Genomic structure of the mouse *Nfe2* gene indicating the three exons and all intron/exon boundaries. Numbers above the exon sequences refer to codons in p45 NF-E2. The first exon (clear box) consists of 5'-untranslated sequences; the second and third exons (shaded boxes) encode a 373-amino-acid product. Note that a non-consensus splice site containing a G-C dinucleotide is present at the exon 1/intron 1 boundary¹⁸. Arrows indicate the positions and orientations of the six oligonucleotides used as PCR primers to subclone the immediate 5' untranslated and coding regions of the *Nfe2* allele from normal and *mk/mk* mice. The sequences of these oligonucleotides are shown below in 5' to 3' orientation; underlined sequences are present within the gene; the 5' terminal sequences contain *Bam*HI or *Eco*RI restriction sites to facilitate subcloning.

1. GACCGGATCCGAGGGCTTCGGAGAGACCGCCGGTG;
2. GACCGAATTCACCTGCCAGCCTGACCTCCCCAG;
3. GACCGGATCCTCTGACTTGGTGGGAGCAACTCTAA;
4. GACCGAATTCAGGCAAGAGCCACCCCTACTCAC;
5. GACCGGATCCACCACTGCCCAACATAGTCGATCTT;
6. GACCGAATTCATCAGAAGGGAATGAGGGAGTCTCT.

in the inbred MK/Re mice. By analysing restriction-fragment length polymorphisms, we established that MK/Re-*mk/mk* mice carry the DBA/2J *Nfe2* allele (data not shown). The immediate 5' untranslated region and coding exons of the *Nfe2* gene were isolated from DNA amplified from both *mk/mk* and DBA/2J mice by polymerase chain reaction (PCR). These fragments were subcloned and sequenced (see Fig. 2 legend for details of primers). Additionally, DNA from both 129/Sv mice and a DBA-derived erythroleukaemia cell line were sequenced. As shown in Fig. 3, a single base change (T→C) at nucleotide 851, which predicts a valine-to-alanine substitution at amino acid 173, was present in all subclones of DNA from MK/Re-*mk/mk* mice. This missense mutation lies near a potential site for phosphorylation by protein kinase A (Ser 170). The valine residue is invariant in wild-type DBA/2J, in the erythroleukaemia cell line, and in 129/Sv DNA, as well as in the human gene encoding p45 NF-E2 (S.H.O. and N.C.A., unpublished data). These findings and the inbred nature of the DBA/2J mouse strain indicate that the observed change in the *mk* allele is not a neutral sequence polymorphism.

Although NF-E2 DNA-binding activity has been demonstrated only in haematopoietic cells, the *mk* phenotype suggests that p45 NF-E2 plays a role in intestinal iron uptake and storage, as well as in regulation of haemoglobin synthesis in erythroid cells. To investigate this, we assayed gene expression by northern blot hybridization using RNA isolated from a variety of normal mouse tissues. As expected, a 1.8-kb transcript was detected in erythropoietic tissues, including spleen and fetal liver, phenylhydrazine-induced adult reticulocytes and uninduced reticulocytes from fetal and newborn mice; no expression was detected in brain, heart, kidney, liver, lung, muscle or testis (Fig. 4a). Initially, no expression was detected in normal mouse duodenum (Fig. 4b, lanes 1–4). But as intestinal expression of p45 NF-E2 mRNA would most likely be restricted to a small subset of cells within the intestine, low hybridization signals would be expected. Reasoning that intestinal expression might be easier to detect under conditions of active iron uptake, we re-examined duodenal mRNA isolated from both normal mice and severely anaemic β -thalassaemic (C57BL/6J-*Hbb*^{d-th3}/*Hbb*^{d-th3}) mice, which show greatly accelerated iron absorption¹⁴, and increased the amount of blotted mRNA. As shown in Fig. 4b, lane 7, p45 NF-E2 mRNA was detected in the duodenum of thalassaemic mice. In addition, there was weak expression in the normal duodenum (lane 5). Expression of p45 NF-E2 mRNA in the intestine does not merely reflect extramedullary haematopoiesis or the presence of circulating erythroid cells, as RNAs for two transcripts that are highly expressed in erythroid precursors



METHODS. p45 NF-E2 sequences were isolated by screening a genomic library, prepared from 129/Sv DNA in the bacteriophage vector λ DashII (Stratagene), with full-length cDNA. The cloned insert was subcloned as an 18-kb *Not*I fragment into a modified form of the vector pSC-3Z. Standard procedures were used for restriction mapping, subcloning and sequencing of intron/exon boundaries¹⁹. Intron sizes were determined by PCR using primers in each exon.

and circulating reticulocytes, GATA-1 and erythroid ankyrin (*Ank-1*)^{15–17}, could not be detected in intestinal RNA by northern blotting (Fig. 4b) or reverse-transcribed PCR (Fig. 4c), respectively. Our results therefore implicate the missense mutation in *Nfe2* as the underlying basis for the intestinal absorptive defect in *mk/mk* mice. The preliminary data presented here suggest that p45 NF-E2 mRNA expression correlates with the iron absorptive state. Additional quantitative studies will be necessary to confirm this observation.

The valine-to-alanine substitution at position 173 lies outside both the basic region DNA-binding domain and the adjacent leucine zipper dimerization domain of p45 NF-E2. Accordingly,

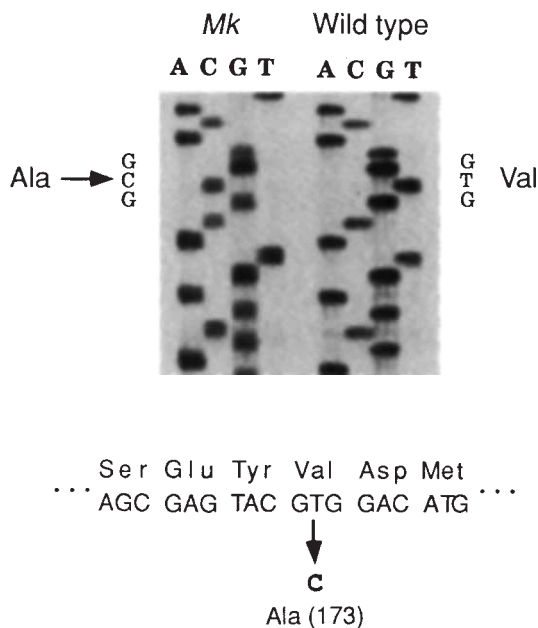


FIG. 3 Missense mutation in the *mk* *Nfe2* allele. A portion of the nucleotide and amino-acid sequences of p45 NF-E2 is shown. Below the wild-type sequence, the single change (T→C) at nucleotide 851 and resulting amino-acid change (Val→Ala) are indicated. Appropriate portions of sequencing gels are also shown.

METHODS. DNA was amplified by PCR from *mk/mk* DNA using primers indicated in Fig. 2. Amplified fragments were subcloned into Bluescript KSII + for double-stranded sequencing using standard methods¹⁹. Four of four *mk* clones contained the indicated mutation.

expression of recombinant p45 NF-E2 protein bearing this mutation in monkey kidney (COS) cells generates NF-E2 DNA binding activity indistinguishable from that generated by expression of wild-type cDNA (not shown). We conclude that the *mk* mutation alters another aspect of NF-E2 function, perhaps its potential to activate transcription. Among other possibilities, the *mk* mutation might influence phosphorylation of the nearby canonical protein kinase A recognition site. Regardless of the mechanism, the phenotype of *mk/mk* mice indicates that the mutation adversely affects the ability of native NF-E2 to regulate globin genes and genes involved in iron metabolism.

The demonstration of a missense mutation within the *mk* gene provides persuasive genetic evidence of an important functional role for p45 NF-E2 *in vivo*. Moreover, our findings suggest a broader role for NF-E2 than its involvement in locus control

region activity and in the regulation of selected erythroid promoters of haem biosynthetic enzymes. Specifically, this nuclear regulator may provide a means by which iron uptake/absorption by the intestine is coupled to the extraordinary demand for iron in developing red cells. The diverse and complementary roles of NF-E2 in both erythroid cells and intestine illustrate how a single regulator may participate at several steps in a common pathway of development in the whole organism.

Note added in proof: We have since sequenced *Nfe2* from the microcytosis mutation on the SEC/1Re strain, and find the identical mutation at amino acid 173. □

Received 1 December 1992; accepted 29 March 1993.

- Mignotte, V., Wall, L., deBoer, E., Grosfeld, F. & Romeo, P.-H. *Nucleic Acids Res.* **17**, 37–54 (1989).
- Philipsen, S., Talbot, D., Fraser, P. & Grosfeld, F. *EMBO J.* **9**, 2159–2178 (1990).
- Ney, P. A., Sorrentino, B. P., McDonagh, K. T. & Nienhuis, A. W. *Genes Dev.* **4**, 993–1006 (1990).
- Jarman, A. P. et al. *Molec. cell Biol.* **11**, 4679–4689 (1991).
- Taketani, S., Inazawa, J., Nakahashi, Y., Abe, T. & Tokunaga, R. *Eur. J. Biochem.* **205**, 217–222 (1992).
- Andrews, N. C., Erdjument-Bromage, H., Davidson, M. B., Tempst, P. & Orkin, S. H. *Nature* **362**, 722–728 (1993).
- Russell, E. S., McFarland, E. C. & Kent, E. L. *Transp. Proc.* **2**, 144–151 (1970).
- Russell, E. S. et al. *Blood* **35**, 838–850 (1970).
- Edwards, J. A. & Hoke, J. E. *Blood* **46**, 381–388 (1975).
- Edwards, J. A. & Hoke, J. E. *Proc. soc. exp. biol. Med.* **141**, 81–84 (1972).
- Bannerman, R. M., Edwards, J. A., Kreimer-Birnbaum, M., McFarland, E. & Russell, E. S. *Br. J. Haemat.* **23**, 235–245 (1972).
- Harrison, D. E. *Blood* **40**, 893–901 (1972).
- Buchanan, G. R. & Sheehan, R. G. *J. Pediatr.* **98**, 723–728 (1981).
- Popp, R. A., Popp, D. M., Johnson, F. M., Skow, L. C. & Lewis, S. E. *Ann. N.Y. Acad. Sci.* **445**, 432–444 (1985).
- Orkin, S. H. *Cell* **63**, 665–672 (1990).
- Tsai, S. F. et al. *Nature* **339**, 446–451 (1989).
- Peters, L. L. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 5749–5753 (1992).
- Fischer, H. D., Dodgson, J. B., Hughes, S. & Engal, J. D. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2733–2737 (1984).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* 2nd edn (Cold Spring Harbor Laboratory Press, New York, 1989).
- Chomczynski, P. & Sacchi, N. *Analyt. Biochem.* **162**, 156–159 (1987).
- Aviv, H. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **69**, 1408–1412 (1972).
- Chui, D. H., Patterson, K. M. & Bayley, S. T. *Br. J. Haemat.* **44**, 431–439 (1980).
- Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **132**, 6–13 (1983).

ACKNOWLEDGEMENTS. We thank L. Washburn for animal husbandry and for advice on *Nfe2* mapping, S. Goff for help in isolating the p45 NF-E2 genomic clone, M. Fleming for discussion, and J. Barker and M. Davison for reviewing the manuscript. Supported by a Merck-AFCR Foundation MD-PhD postdoctoral fellowship (N.C.A.) and NIH grants (S.E.L., E.M.E. and S.H.O.). S.H.O. is an investigator of The Howard Hughes Medical Institute.

Projection structure of rhodopsin

Gebhard F. X. Schertler, Claudio Villa & Richard Henderson

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

LIGHT absorption by the visual pigment rhodopsin^{1,2} triggers, through G-protein coupling, a cascade of events in the outer segment of the rod cell of the vertebrate retina that results in membrane hyperpolarization and nerve excitation^{3–5}. Rhodopsin, which contains 348 amino acids^{6–8}, has seven helices that cross the disk membrane^{6,9} and its amino terminus is extracellular. A wealth of biochemical data is available for rhodopsin: 11-*cis* retinal is bound¹⁰ to lysine 296 in helix VII; glutamic acid 113 on helix III is the counterion to the protonated Schiff's base^{11,12}; a disulphide bridge, cysteine 110–187, connects helix III to the second extracellular loop e2 (refs 13, 14); the carboxy terminus has two palmitoylated cysteines forming a cytoplasmic loop i4 (ref. 15); three intracellular loops i2, i3 and i4 mediate activation of the heterotrimeric G protein transducin^{16,17}; glutamic acid 135 and arginine 136 at the cytoplasmic end of helix III affect binding of transducin¹⁸. But to provide a framework to interpret these data, not only for rhodopsin but for other G-protein-coupled receptors, requires the structure to be determined. Here we present a projection map of rhodopsin showing the configuration of the helices.

We have reconstituted purified bovine rhodopsin¹⁹ into lipid bilayers to obtain two-dimensional crystals²⁰. In our best samples, 10% of the vesicles contained two-dimensional arrays

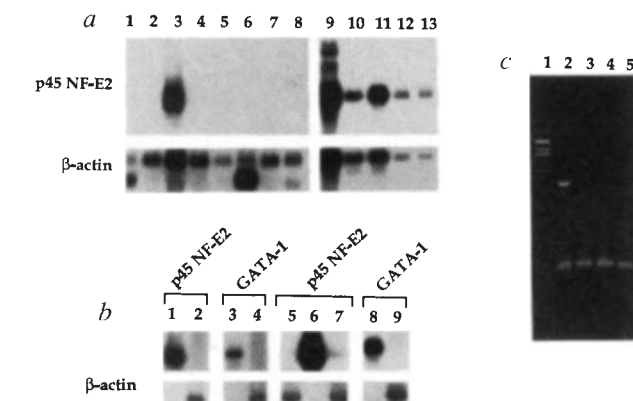


FIG. 4 *a*, Survey of p45 NF-E2 expression in mouse tissues. Lanes 1–8 contain 2 µg of mRNA. Lane 1, heart; lane 2, brain; lane 3, spleen; lane 4, lung; lane 5, liver; lane 6, skeletal muscle; lane 7, kidney; lane 8, testis; lane 9, 5 µg phenylhydrazine-induced spleen mRNA; lane 10, 5 µg phenylhydrazine-induced reticulocyte mRNA; lane 11, 1 µg fetal liver mRNA; lane 12, 1 µg fetal reticulocyte mRNA; lane 13, 1 µg newborn reticulocyte mRNA. p45 NF-E2 mRNA is ~1.8 kb (relative to 18S and 28S ribosomal RNA). The origin of the additional transcripts detected in phenylhydrazine-induced spleen mRNA (lane 3) is unknown, but the mRNAs are of the sizes predicted for nuclear precursor RNAs. *b*, Expression of p45 NF-E2 in the duodenum. Lanes 1, 3, 6 and 8, 0.5 µg phenylhydrazine-induced reticulocyte mRNA; lanes 2 and 4, 1 µg normal duodenum mRNA; lane 5, 2.5 µg normal duodenum mRNA; lanes 7 and 9, 2.5 µg β-thalassaemic (C57BL/6J-*Hbb*^{d-th3}/*Hbb*^{d-th3}) mRNA. The hybridization probe (p45 NF-E2 or GATA-1) is shown at top. mRNA samples in lanes 2 and 4 and lanes 7 and 9 are from the same preparations. Lanes 1–4 were exposed for 60 h, lanes 5–9 for 24 h. *c*, RT-PCR amplification using primers specific for mouse erythroid ankyrin (*Ank-1*). Lane 1, pUC19 *M*_r markers; lane 2, normal reticulocyte; lane 3, β-thalassaemic duodenum; lane 4, normal duodenum; lane 5, negative control.

METHODS. Total cellular RNA was prepared according to ref. 20. Polyadenylated RNA was isolated by oligo(dT)–cellulose chromatography²¹. The left blot in *a* was obtained from Clontech (Palo Alto). Phenylhydrazine treatment of adult mice induced peripheral reticulocytosis to >95% and erythroid hyperplasia of the spleen²². To isolate intestinal RNA, animals were anaesthetized with tribromoethanol and perfused through the left ventricle with 20 ml sterile phosphate-buffered saline. Duodenum refers to a segment of 3–5 cm beginning at the pylorus of the stomach. For northern blotting, samples were fractionated in agarose–formaldehyde gels and transferred to nylon filters (Zetabind). The p45 NF-E2 probe (1.5 kb) and the GATA-1 probe (1.3 kb) were labelled by the random hexamer method to specific activities of 4 × 10⁹ c.p.m. per µg (ref. 23). Hybridization conditions were as specified for Zetabind filters with 1% SDS. Filters were washed at high stringency (0.1 × SSC, 0.1% SDS at 65 °C) and exposed to Kodak XAR-5 film at –70 °C with two intensifying screens. For RT-PCR, conditions were as specified for the GeneAmp kit (Perkin–Elmer Cetus). Total RNA (10 µg) was reverse-transcribed and amplified using primers specific for mouse *Ank-1* (forward primer, CGTGAGCTGGTGGTCTGAGGAGTG; reverse primer, TCTTGCCACTTGTGTCATCCGTCAT), which yields an 800-bp fragment. Thalassaemic duodenal RNA in the absence of reverse transcriptase was used as a negative control.

of rhodopsin (Fig. 1a). In most cases, both the top and bottom layers of the vesicle were crystalline. Optical diffraction indicated an orthorhombic lattice with $a = 43 \text{ \AA}$ and $b = 140 \text{ \AA}$. Systematic absence of reflections suggested the presence of 2-fold screw axes along the b -axis. Examination of computer-averaged maps showed 2-fold axes and 2-fold screw axes parallel to the a - and b -axes, respectively, and a 2-fold axis perpendicular to the membrane. This was confirmed by examination of the phase relationships between the reciprocal space Fourier components. The space group of the crystals was thus shown

to be $p222_1$, which is different from previous two-dimensional crystals of rhodopsin that have been reported by other investigators²¹⁻²³.

The crystals were further analysed unstained as frozen hydrated specimens (Fig. 1b). The data from 13 crystalline areas described in Table 1 were merged using an approximate contrast transfer function correction and brought to the same amplitude scale. The phase origin and defocus of individual images were then refined iteratively. For each reflection, the amplitude was corrected for the contrast transfer function and the structure

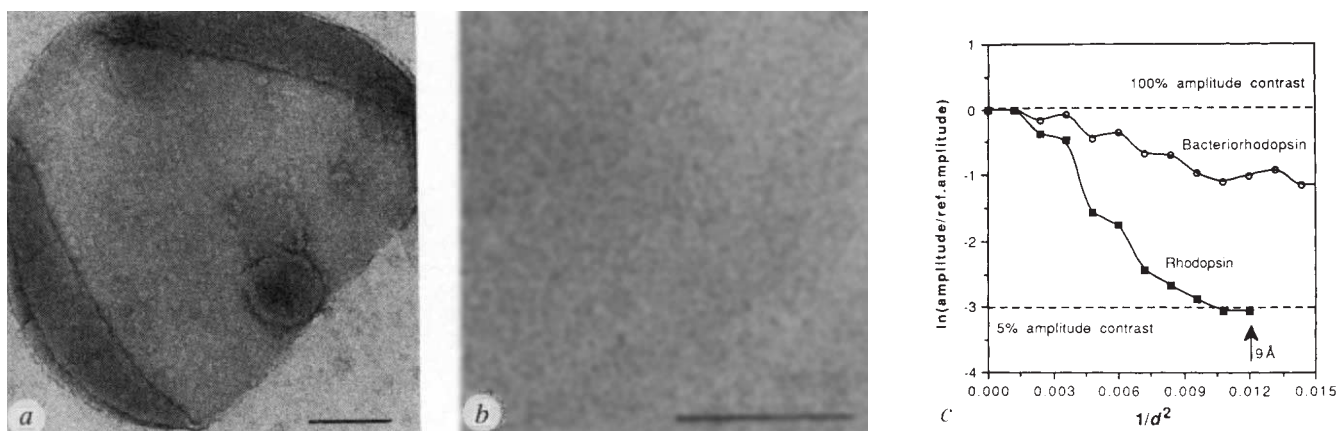
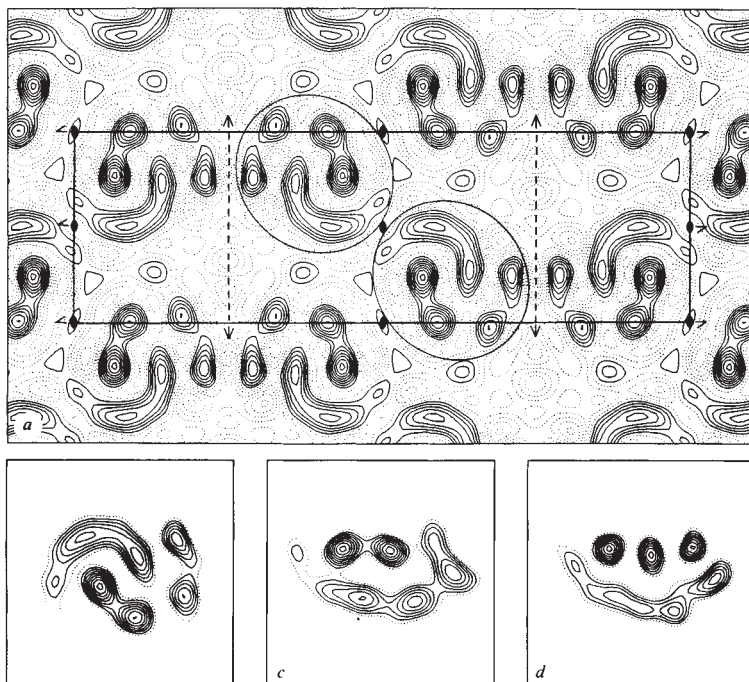


FIG. 1 Two-dimensional crystals of bovine rhodopsin. We purified rhodopsin from bovine rod outer segments solubilized in lauryldimethylamine oxide, using a concanavalin-A column¹⁹ followed by Sephadex G50 gel filtration and ion-exchange chromatography (S-Sepharose fast flow, Pharmacia) in n -octyl tetraoxy-ethylene (C_8E_4). The purified rhodopsin was then reconstituted in soybean phosphatidylcholine (Sigma) at low lipid-to-protein ratios (10–50 lipids per rhodopsin molecule in 20 mM HEPES, 0.1 M NaCl, 10 mM $MgCl_2$, 1 mM dithiothreitol (DTT), 3 mM NaN_3 , pH 7.0 at 20 °C)²⁰. The detergent was removed by dialysis. The samples were screened for crystallization by electron microscopy using uranyl acetate as a negative stain. Two-dimensional crystals formed after 5–20 days, but only from freshly prepared rhodopsin. No change in the extent of crystallinity occurred over a period of several months. *a*, Electron micrograph of a two-dimensional crystal of rhodopsin stained with 1% uranyl acetate. In most vesicles, two crystalline layers representing the two sides are found on top of each other. The two lattices (orthorhombic, $p222_1$, $a = 43 \text{ \AA}$, $b = 140 \text{ \AA}$) can easily be treated separately in the Fourier transform of the images. *b*, Electron micrograph of a frozen hydrated rhodopsin crystal. A drop of the crystalline suspension was applied to a carbon-coated electron microscope grid that had been glow-discharged in the presence of amylamine. The surplus suspension was blotted off and the specimen was rapidly frozen in liquid ethane. All operations, including the freezing, were done in dim red light ($>650 \text{ nm}$). The vitrified sample was either stored in liquid nitrogen or immediately transferred at -160 °C to a cryo stage (Gatan) and then to the electron microscope (Philips EM420) equipped with a twin-bladed anti-contaminator. Vesicles were selected by searching with a dim, defocused beam in diffraction mode. The microscope was focused on an area near to the vesicle at high magnification using a Philips low-dose unit. Electron micrographs were recorded at 120 kV, with a magnification of $\times 36,000$ and underfocus from 5,200 to 16,000 $\text{\AA}$, using an electron dose of $5\text{--}10 \text{ e}^-/\text{\AA}^2$. Scale bars, $0.1 \text{ }\mu\text{m}$. The crystals were further analysed unstained as frozen hydrated specimens (*b*). The 8 best pictures, which included 13 independent crystalline areas, were chosen by optical diffraction from several hundred micrographs for further processing. They all showed clearly the (1, 1), (1, 2), (1, 3), (0, 4) and (0, 8) reflections but nothing was visible beyond $15 \text{ \AA}$. The best areas of each image were selected and digitized ($2,000 \times 2,000$ pixel at $10 \text{ }\mu\text{m}$ intervals; Joyce Loebel Mk4 densitometer). The images were corrected for distortions in the crystal lattice using a 200×200 pixel reference area²⁶. For a second pass of correction, a 150×150 pixel area from the filtered image of the initially corrected area was used as the reference. Finally, the area representing the extent of the crystal was boxed off before calculating the amplitudes and phases from each crystal. *c*, Resolution-dependent fading of image amplitudes for two-dimensional crystals of rhodopsin and bacterio-

rhodopsin. The natural logarithm of the ratio between average image amplitudes and average bacteriorhodopsin electron diffraction amplitudes^{24,27} in resolution zones against $1/d^2$ was plotted. The image amplitudes of the rhodopsin crystals fade more rapidly because of the limited order of the crystals. At $9 \text{ \AA}$ resolution, the amplitudes of rhodopsin have 5% of the theoretically expected amplitude contrast. The image amplitudes were rescaled as a function of resolution, to compensate for the loss of contrast, using a correction derived from the curve of the rhodopsin crystals. *d*, All unique Fourier components measured to $7 \text{ \AA}$ resolution are shown in reciprocal space. The size of the boxes indicates the phase error (before rounding to 0° or 180°) associated with each measurement ($1 < 8^\circ$, $2 < 14^\circ$, $3 < 20^\circ$, $4 < 25^\circ$, $5 < 32^\circ$, $6 < 39^\circ$, $7 < 45^\circ$, $8 < 90^\circ$; values from 1–4 are given as numbers in the boxes, values from 5–8 are indicated by decreasing box size). In this case, 90° is random. Only data to $9 \text{ \AA}$ resolution, indicated by the circle, were used for calculating the projection map.

FIG. 2 *a*, Projection density map of rhodopsin at 9 Å resolution, calculated from merged and corrected image amplitudes and phases obtained from 13 independent crystalline areas. A unit cell with the *a*-axis vertical and the *b*-axis horizontal is shown. The space group is $p222_1$ ($a=43$ Å, $b=140$ Å). The 2-fold axes perpendicular and parallel to the membrane plane and the 2-fold screw axes along *b* are indicated. One unit cell contains four rhodopsin molecules, two of which have been circled. Zero and negative contours are shown as dotted lines. *b*, The density for a single rhodopsin molecule at 9 Å resolution is shown, negative contours were omitted and the zero contour is represented by a dotted line. *c*, The density for a single bacteriorhodopsin molecule at 9 Å resolution is shown using the same contour levels and the same scale. *d*, The density for a single bacteriorhodopsin²⁸ molecule at 7 Å resolution is shown in the same way.



factors were averaged vectorially. Final phases were obtained by rounding to either 0° or 180° because the projection is centrosymmetric. The image amplitudes were rescaled as a function of resolution by applying a correction derived from a comparison with the average diffraction amplitude of bacteriorhodopsin at a given resolution, as discussed in Fig. 1c. The data were significant to 8.7 Å but were weaker and therefore noisier at higher resolution. The quality and resolution of the amplitude and phase data are shown in Table 1 and Fig. 1c, d.

We have chosen to calculate a projection at 9 Å resolution as shown in Fig. 2. The projection density map shows, within the region occupied by each rhodopsin molecule, an elongated arc-shaped feature and four resolved peaks of density. We interpret the four peaks to represent four transmembrane helices oriented nearly perpendicular to the membrane. The elongated arc-shaped feature probably represents the three remaining, more tilted helices. The structure of rhodopsin is therefore a bundle of seven transmembrane helices oriented mostly perpendicular to the membrane plane.

In Fig. 2b the projection density for a single rhodopsin molecule can be compared with the density of bacteriorhodopsin²⁴ shown at 9 Å in Fig. 2c and at 7 Å resolution in Fig. 2d. The projection structure of rhodopsin is less elongated and slightly wider and the helices are tilted differently from bacteriorhodopsin.

Using additional information, an attempt has been made to

TABLE 1 Image statistics

Number of images	8
Number of crystalline areas	13
Variation in cell parameters	±3%
Range of defocus	5,200–16,000 Å
Range of astigmatism	0–6,400
Image tilt	0–8°
Total number of observations	1,755 reflections*
Total number of unique observations	109 reflections*
Overall phase residual	29.7°*†

Resolution range in Å	Number of unique reflections	Phase residual†
200–14	27	16.8°
14–10	28	27.2°
10–8.7	15	29.3°
8.7–7.0	39	44.3°

* For all data to 7 Å resolution.

† Compared to 0°/180°; 45° is a random value.

assign the seven hydrophobic stretches in the polypeptide sequence to the peaks in this projection structure²⁵. But a more certain interpretation will be possible once the collection of data from tilted specimens allows calculation of a three-dimensional map so that the helix tilt angles can be seen. □

Received 20 November 1992; accepted 5 February 1993.

- Hargrave, P. A. *Curr. Opin. struct. Biol.* **1**, 575–581 (1991).
- Henderson, R. & Schertler, G. F. X. *Phil. Trans. R. Soc. B* **326**, 379–389 (1990).
- De Grip, W. J. *Photochem. Photobiol.* **48**, 799–810 (1988).
- Chabre, M. & Deterre P. *Eur. J. Biochem.* **179**, 255–266 (1989).
- Stryer, L. *J. biol. Chem.* **266**, 10711–10714 (1991).
- Ovchinnikov, Y. A. *et al. Bioorg. Khim.* **8**, 1424–1427 (1982).
- Hargrave, P. A. *et al. Biophys. struct. Mech.* **9**, 235–244 (1983).
- Nathans, J. & Hogness, D. S. *Cell* **34**, 807–814 (1983).
- Khorana, H. G. *J. biol. Chem.* **267**, 1–4 (1992).
- Bownds, D. *Nature* **216**, 1178–1181 (1967).
- Zhukovskiy, E. A. & Oprian, D. D. *Science* **246**, 928–930 (1989).
- Sakmar, T. P., Franke, R. R. & Khorana, H. G. *Proc. natn. Acad. Sci. U.S.A.* **86**, 8309–8313 (1989).
- Karnik, S. S., Sakmar, T. P., Chen, H.-B. & Khorana, H. G. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8459–8463 (1988).
- Karnik, S. S. & Khorana, H. G. *J. biol. Chem.* **265**, 17520–17524 (1990).
- Ovchinnikov, Y. A., Abdulaev, N. G. & Bogachuk, A. S. *FEBS Lett.* **230**, 1–5 (1988).
- Hargrave, P. A., Hamm, H. E. & Hofmann, K. P. *Bioessays* (in the press).

- König, B. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 6878–6882 (1989).
- Franke, R. R., König, B., Sakmar, T. P., Khorana, H. G. & Hofmann, K. P. *Science* **250**, 123–125 (1990).
- De Grip, W. J. *Meth. Enzym.* **81**, 197–207 (1982).
- Kühlbrandt, W. *Q. Rev. Biophys.* **25**, 1–49 (1992).
- Corless, J. M., McCaslin, D. R. & Scott, B. L. *Proc. natn. Acad. Sci. U.S.A.* **79**, 1116–1120 (1982).
- Dratz, E. A., Van Breemen, J. F. L., Kampe, K. M. P., Keegstra, W. & Van Bruggen, E. F. J. *Biochim. biophys. Acta* **832**, 337–342 (1985).
- Demin, V. V., Yurkova, E. V., Kuzin, A. P., Barnakov, A. N. & Abdulaev, N. G. in *Retinal Proteins* (ed. Ovchinnikov, Y. A.) 519–524 (VNU Science, Utrecht, 1987).
- Henderson, R. *et al. molec. Biol.* **213**, 899–929 (1990).
- Baldwin, J. M. *EMBO J.* (in the press).
- Henderson, R., Baldwin, J. M., Downing, K. H., Lepault, J. & Zemlin, F. *Ultramicroscopy* **19**, 147–178 (1986).
- Ceska, T. A. & Henderson, R. *J. molec. Biol.* **213**, 539–560 (1990).
- Unwin, P. N. T. & Henderson, R. *J. molec. Biol.* **94**, 425–440 (1975).

ACKNOWLEDGEMENTS. We thank J. Findlay and H. Saibil for their advice in the early stages of the project, R. Lucke, C. Raeburn and R. Blows for building equipment, and EMBO for a fellowship (G.F.X.S.).